DOCUMENT RESUME

ED 340 053 CS 507 639

AUTHOR	Studdert-Kennedy, Michael, Ed.
TITLE	Status Report on Speech Research, January-June 1991.
INSTITUTION	Haskins Labs., New Haven, Conn.
SPONS AGENCY	National Inst. of Child Health and Human Development (NIH), Bethesda, MD.; National Inst. of Health (DHHS), Bethesda, MD. Biomedical Research Support Grant Program.; National Inst. on Deafness and Other Communications Disorders, Bethesda, MD.; National Science Foundation, Washington, D.C.
REPORT NO	SR-105/106
PUB DATE	Jul 91
CONTRACT	NO1-HD-5-2910
NOTE	290p.; For previous report (July-December 1990), see ED 331 100.
PUB TYPE	Reports - Research/Technical (143) -- Collected Works - General (020)
EDRS PRICE	MF01/PC12 Plus Postage.
DESCRIPTORS	Beginning Reading; Communication Research; *Comprehension; Elementary Secondary Education; Higher Education; *Language Processing; *Memory; Music Techniques; *Phonology; *Reading Difficulties; *Speech Communication; Stuttering
IDENTIFIERS	Speech Research

ABSTRACT

One of a series of semiannual reports, this publication contains 18 articles which report the status and progress of studies on the nature of speech, instrumentation for its investigation, and practical applications. Articles and their authors are as follows: "Phonology and Beginning Reading Revisited (Isabelle Y. Liberman); "The Role of Working Memory in Reading Disability" (Susan Brady); "Working Memory and Comprehension of Spoken Sentences: Investigations of Children with Reading Disorder" (Stephen Crain and others); "Explaining Failures in Spoken Language Comprehension by Children with Reading Disability" (Stephen Crain and Donald Shankweiler); "How Early Phonological Development Might Set the Stage for Phoneme Awareness" (Anne E. Fowler); "Modularity and the Effects of Experience" (Alvin M. Liberman and Ignatius G. Mattingly); "Modularity and Disassociations in Memory Systems" (Robert G. Crowder); "Representation and Reality: Physical Systems and Phonological Structure" (Catherine P. Browman and Louis Goldstein); "Young Infants' Perception of Liquid Coarticulatory Influences on Following Stop Consonants" (Carol A. Fowler and others); "Extracting Dynamic Parameters from Speech Movement Data" (Caroline L. Smith and others); "Phonological Underspecification and Speech Motor Organization" (Suzanne E. Boyce and others); "Task Coordination in Human Prehension" (Patrick Haggard); "Masking and Stimulus Intensity Effects on Duplex Perception: A Confirmation of the Dissociation between Speech and Nonspeech Modes" (Shlomo Bentin and Virginia Mann); "The Influence of Spectral Prominence on Perceived Vowel Quality" (Patrice Speeter Beddor and Sarah Hawkins); "On the Perception of Speech from Time-Varying Acoustic Information: Contributions of Amplitude Variation" (Robert E. Remez and Philip E. Rubin); "Subject Definition and Selection Criteria for Stuttering Research in Adult Subjects" (Peter J. Alfonso); "Vocal Fundamental Frequency Variability in Young Children: A Comment on 'Developmental Trends in Vocal Fundamental Frequency of Young Children' by M. Robb and J. Saxman" (Margaret Lahey and others); and "Patterns of Expressive Timing in Performances of a Beethoven Minuet by Nineteen Famous Pianists" (Bruno H. Repp). An appendix lists DTIC and ERIC numbers for status reports in this series since 1990. (PRA)

ED340053

Haskins Laboratories Status Report on

Speech Research

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

A. Dadourian

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

AS507639

SR-105/106
JANUARY-JUNE 1991

BEST COPY AVAILABLE

Haskins Laboratories Status Report on

Speech Research

SR-105/106
JANUARY-JUNE 1991

NEW HAVEN, CONNECTICUT

Distribution Statement

Editor
Michael Studdert-Kennedy

Production Staff
Yvonne Manning
Zefang Wang

This publication reports the status and progress of studies on the nature of speech, instrumentation for its investigation, and practical applications.

Distribution of this document is unlimited. The information in this document is available to the general public. Haskins Laboratories distributes it primarily for library use. Copies are available from the National Technical Information Service or the ERIC Document Reproduction Service. See the Appendix for order numbers of previous Status Reports.

Correspondence concerning this report should be addressed to the Editor at the address below:

Haskins Laboratories
270 Crown Street
New Haven, Connecticut
06511-6695

Phone: (203) 865-6163 *FAX:* (203) 865-8963 *Bitnet:* HASKINS@YALEHASK
Internet: HASKINS%YALEHASK@VENUS.YCC.YALE.EDU

This Report was reproduced on recycled paper

Acknowledgment

The research reported here was made possible in part by support from the following sources:

National Institute of Child Health and Human Development

Grant HD-01994
Grant HD-21888
Contract NO1-HD-5-2910

National Institute of Health

Biomedical Research Support Grant RR-05596

National Science Foundation

Grant BNS-8820099

National Institute on Deafness and Other Communication Disorders

Grant DC 00121
Grant DC 00183
Grant DC 00403
Grant DC 00016
Grant DC 00594

Investigators

Arthur Abramson*
Peter J. Alfonso*
Thomas Baer*
Eric Bateson*
Fredericka Bell-Berti*
Catherine T. Best*
Susan Brady*
Catherine P. Browman
Franklin S. Cooper*
Stephen Crain*
Robert Crowder*
Lois G. Dreyer*
Alice Faber
Laurie B. Feldman*
Janet Fodor*
Carol A. Fowler*
Louis Goldstein*
Carol Gracco
Vincent Gracco
Vicki L. Hanson*
Katherine S. Harris*
John Hogden
Leonard Katz*
Rena Arens Krakow*
Andrea G. Levitt*
Alvin M. Liberman*
Diane Lillo-Martin*
Leigh Lisker*
Anders Löfqvist*
Virginia H. Mann*
Ignatius G. Mattingly*
Nancy S. McGarr*
Richard S. McGowan
Patrick W. Nye
Kiyoshi Oshima†
Lawrence J. Raphael*
Bruno H. Repp
Philip E. Rubin
Elliot Saltzman
Donald Shankweiler*
Michael Studdert-Kennedy*
Michael T. Turvey*
Douglas Whalen

*Part-time
†Visiting from University of Tokyo, Japan

Technical/Administrative Staff

Philip Chagnon
Alice Dadourian
Michael D'Angelo
Betty J. DeLise
Lisa Fresa
Vincent Gulisano
Donald Hailey
Maura Herlihy
Raymond C. Huey*
Marion MacEachron*
Yvonne Manning
Joan Martinez
William P. Scully
Richard S. Sharkany
Zefang Wang
Edward R. Wiley

Students*

Melanie Campbell
Sandra Chiang
Margaret Hall Dunn
Terri Erwin
Elizabeth Goodell
Joseph Kalinowski
Laura Koenig
Betty Kollia
Simon Levy
Salvatore Miranda
Maria Mody
Weijia Ni
Mira Peter
Nian-qi Ren
Christine Romano
Joaquin Romero
Arlyne Russo
Jeffrey Shaw
Caroline Smith
Mark Tiede
Qi Wang
Yi Xu
Elizabeth Zsiga

Contents

Phonology and Beginning Reading Revisited
Isabelle Y. Liberman 1

The Role of Working Memory in Reading Disability
Susan Brady 9

Working Memory and Comprehension of Spoken Sentences: Investigations of Children with Reading Disorder
Stephen Crain, Donald Shankweiler, Paul Macaruso, and Eva Bar-Shalom 23

Explaining Failures in Spoken Language Comprehension by Children with Reading Disability
Stephen Crain and Donald Shankweiler 43

How Early Phonological Development Might Set the Stage for Phoneme Awareness
Anne E. Fowler 53

Modularity and the Effects of Experience
Alvin M. Liberman and Ignatius G. Mattingly 65

Modularity and Dissociations in Memory Systems
Robert G. Crowder 69

Representation and Reality: Physical Systems and Phonological Structure
Catherine P. Browman and Louis Goldstein 83

Young Infants' Perception of Liquid Coarticulatory Influences on Following Stop Consonants
Carol A. Fowler, Catherine T. Best, and Gerald W. McRoberts 93

Extracting Dynamic Parameters from Speech Movement Data
Caroline L. Smith, Catherine P. Browman, Richard S. McGowan, and Bruce Kay 107

Phonological Underspecification and Speech Motor Organization
Suzanne E. Boyce, Rena A. Krakow, and Fredericka Bell-Berti 141

Task Coordination in Human Prehension
Patrick Haggard 153

Masking and Stimulus Intensity Effects on Duplex Perception: A Confirmation of the Dissociation Between Speech and Nonspeech Modes
Shlomo Bentin and Virginia Mann 173

The Influence of Spectral Prominence on Perceived Vowel Quality
Patrice Speeter Beddor and Sarah Hawkins 187

On the Perception of Speech from Time-varying Acoustic Information: Contributions of Amplitude Variation
Robert E. Remez and Philip E. Rubin 215

7

Subject Definition and Selection Criteria for Stuttering Research in Adult Subjects
Peter J. Alfonso 231

Vocal Fundamental Frequency Variability in Young Children: A Comment on *Developmental Trends in Vocal Fundamental Frequency of Young Children* by M. Robb and J. Saxman
Margaret Lahey, Judy Flax, Katherine Harris, and Arthur Boothroyd 243

Patterns of Expressive Timing in Performances of a Beethoven Minuet by Nineteen Famous Pianists
Bruno H. Repp 247

Appendix 273

Haskins
Laboratories
Status Report on

Speech Research

This issue of the Haskins Laboratories Status Report is dedicated to the memory of Isabelle Y. Liberman (1918-1990) whose theoretical and experimental contributions to the study of reading inspired and guided researchers not only here at Haskins, but in many laboratories across this country and Europe. One of her last papers appears in this issue.

Haskins Laboratories Status Report on Speech Research
1991, SR-105/106, 1-8

Phonology and Beginning Reading Revisited*

Isabelle Y. Liberman

The research of my colleagues and me has, for many years, been guided by the assumption that most problems in learning to read and write stem from deficits in the language faculty, not from deficiencies of a more generally cognitive or perceptual sort. A paper by Alvin Liberman (1988) says in detail how and why we were initially led to that assumption. My aim here is rather to describe the assumption itself, offer data in support, and finally to develop the implications for the teacher and clinician.

A CONTRAST BETWEEN LISTENING AND READING

But first, I should offer some background (Liberman, 1987). To that end, I will consider a few facts about words: how they are produced and perceived, and how differently they are processed in spoken and written language. All words are, of course, formed of combinations and permutations of phonological elements called consonants and vowels. The obvious advantage of forming words in this way is that by using no more than two or three dozen different elements, we can and do produce a large and vastly expandable vocabulary numbering tens of thousands of words. If, on the other hand, each word had to be uniquely and holistically different from every other word, the number we could produce would be limited to the number of different individual signals—sounds, if you will—that a person can efficiently make and perceive; that number is, of course, exceedingly small.

An alphabetic writing system—the one we're concerned with—represents the same string of phonological segments—consonants and vowels—that we use in speaking, the string that distinguishes one word from all others. Then why should it be so hard for many beginning readers to grasp the alphabetic principle? Why can they not quickly begin to read and write as well and as easily as they can already speak and listen? The exact nature of the difficulty has been developed in greater detail elsewhere (A. M. Liberman, 1982, 1988; A. M. Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). I should nevertheless summarize the argument here.

Consider how you and I produce a word like "bag," or more to the point, how we do not produce it. We do not say B A G; we say "bag." That is, we fold three phonological segments—two consonants and one vowel—into a single segment of sound. This we do by a process called "coarticulation." In the case of "bag," we overlap the lip movement appropriate for the initial consonant B with the tongue movement appropriate for the medial vowel A, and then smoothly merge that with the tongue movement appropriate for the final consonant G. Such coarticulation, it should be emphasized, is not careless speech. It is the very essence of speech, the only basis on which phonological structures can be produced at the rapid rates that make words, phrases, and sentences feasible.

Consider now the consequences for perception. When one examines a schematic spectrogram sufficient to produce the word "bag" (A. M. Liberman, 1970), one sees that the three phonologic segments are thoroughly merged in the sound stream. The vowel is not limited to a medial position, but covers the entire length of the syllable. Information about the initial consonant continues well beyond the middle of the signal. Moreover, the center portion of the acoustic signal is providing information not just about the vowel, but about all three perceptual segments at once.

Now we must ask how a listener knows that the word is "bag" and not "tag" or "big" or "bat." The answer is that all this is managed by a biologically coherent system, specifically adapted to the

Preparation of this paper was aided by grants to Haskins Laboratories (NIH-NICHD-HD-01994) and to Yale University/Haskins Laboratories (NIH-21888-01A1).

production and perception of phonetic structures. In production, the specialization automatically converts the phonological representation of the word into the coarticulated movements that convey it. In perception, this specialization effectively runs the process in reverse: it automatically parses the sound so as to recover the coarticulated gestures—that is, the segmented phonetic structure—that caused it.

But notice that all of this is automatic, carried out below the level of conscious awareness. Thus, to say "bag," speakers need not know how it is spelled—that is, what sequence of consonants and vowels it comprises. They have only to think of the word. The phonetic specialization in effect spells it for them. Matters are correspondingly automatic for listeners: on hearing the sound "bag," they need not consciously analyze it into its three constituent elements. For, again, the phonetic specialization does it all, recovering the segments and matching them to the word stored in the lexicon; the listeners are none the wiser about the very complex process that has been carried out on their behalf.

The relation of all this to reading and writing has long seemed to us quite obvious (Liberman, 1971). For if readers and writers are to use the alphabetic principle productively—that is, if they are to deal with words they have never seen in print—they must be quite consciously aware of the phonological structure the letters represent. But nothing in their normal linguistic experience has prepared them for this. Never have the processes by which they normally speak and listen revealed to them that words have internal phonologic structures, and never before have they been in a situation which required them to know that such structures do, in fact, exist.

What then are beginners to make of the three letters that form the word *bag*? If we tell them the letters represent sounds, then we are misleading them, because they know, and we should know, that there are *three* letters but really only *one* sound. If we tell them the letters represent abstract and specifically phonologic gestures, then we are conveying the true state of affairs, but, as a practical matter, that explanation is not likely to help them read.

These considerations led us to several hypotheses. The first was that awareness of phonological structure might be a problem for preliterate children. The second was that individual differences in this ability might be related to success in reading. A third hypothesis was that training in phonological awareness would have a positive effect on reading achievement. The fourth hypothesis was that the weakness in phonological awareness displayed by beginners who have difficulties in learning to read might reflect a more general deficiency in the biological specialization that processes phonological structure in speech. Now I turn to the relevant data.

AWARENESS OF PHONOLOGICAL STRUCTURE IN PRELITERATES

What data do we have about awareness? Some 15 years ago we began to examine developmental trends in phonological awareness by testing the ability of young children to segment words into their constituent elements (Liberman, Shankweiler, Fischer & Carter, 1974). We found that normal preschool children performed rather poorly. We learned further that of the two types of sublexical units—the syllable and the phoneme—the phoneme, which happens to be the unit represented by our alphabet, presented the greater difficulty by far. About half the four- and five-year-olds could segment by syllable, but none of the four-year-olds and only 17 of the five-year-olds could manage segmentation by phoneme. Even at six there was a difference—10 failed on the syllables but 30 failed on the phonemes.

The answer to our first hypothesis is clear from these results and those of studies by many other investigators (see Stanovich, 1982 for a review): Awareness of the phonemic segment, the basic unit of the alphabetic writing system, is indeed hard to achieve, harder than awareness of syllables, and it develops later, if at all. It was also apparent that, like any cognitive achievement, it develops to varying degrees at varying rates in different children. Moreover, a large number of children have not attained awareness of either level of linguistic structure—syllable or phoneme—even at the end of a full year in school. If linguistic awareness does indeed provide entry into the alphabetic system, as we think it does, then these children are the ones we need to worry about.

AWARENESS OF PHONOLOGICAL STRUCTURE AND LITERACY

Much evidence is now available to support our second hypothesis, namely, that awareness of the phonological constituents of words is important for the acquisition of reading. The evidence comes from numerous studies here and abroad. These have all shown that phonological awareness is significantly related to reading success in young

children. In English, there are, to name only a few, studies by Liberman (1973); Goldstein (1976); Fox and Routh (1980); Treiman and Baron (1981); Blachman (1984); Bradley and Bryant (1983); Mann & Liberman (1984) and Olson (1988). All these findings have been supported and indeed extended in Swedish by the carefully controlled, pioneering studies of the correlates of reading disability carried out by Lundberg and his associates in Umea (1980, 1988) and also by the recent studies by Magnusson and Naucler in Lund (1987). In Spanish we cite work by de Manrique and Gramigna (1984), in French by Bertelson's Belgian laboratory (Bertelson, 1988; Morais, Cluytens, and Alegria, 1984),and in Italian by Cossu and associates (Cossu, Shankweiler, Liberman, Tola, & Katz, 1988).

THE EFFECT OF TRAINING IN PHONOLOGICAL AWARENESS

Thus there is now a great deal of evidence to support the hypothesis that deficiency in phonological awareness is related to success in reading. The evidence comes, as we have seen, from studies covering a wide range of ages, many language communities, and a variety of cultural and economic backgrounds, ranging from inner city and rural poor to suburban affluent.

It is of special interest, then, to find that phonological awareness can be taught even in preschool (Ball & Blachman, in press; Content, Morais, Alegria, & Bertelson, 1982; Lundberg, 1988; Lundberg, Frost, & Petersen, 1988; Oloffson & Lundberg, 1983; Vellutino & Scanlon, 1987). Early evidence for the value of such training cames from a landmark study by Bradley & Bryant (1983). In the first of a pair of experiments, they found high correlations between preschoolers' phonological awareness as measured by rhyming tasks and the children's reading and spelling scores several years later. In the second experiment, they found that children with initially low levels of phonological awareness who were trained in the phonological classification of words were later superior in reading and spelling to groups who had had semantic classification training or no training at all. Those trained, in addition, to associate letters with the phonemes were even more successful. New evidence for the positive effects of phonological training on reading achievement has recently been reported by Lundberg (1988) and by Vellutino (Vellutino and Scanlon, 1987).

THE SOURCE OF INDIVIDUAL DIFFERENCES IN AWARENESS

What the research data have shown thus far can be summarized by four major points. The first is that despite adequate speech, preliterate children and adults are not necessarily aware that words have an internal phonological structure. Since the alphabet represents that structure, they are therefore not in a position to use the alphabetic principle. The second is that there are individual differences in the ease with which children become aware of phonological structure. Third, these differences correlate with success in learning to read. And finally, explicit training in the analysis of phonological structure produces not only better speech analyzers but also better readers.

Now we should ask whence comes the abnormally great difficulty that some children have in developing the awareness? There are two possibilities: the problem could reflect difficulty with any cognitive tasks that require analytic ability or, alternatively, it could point to a deficiency in the phonological processor that causes it to set up phonological structures weakly. In the latter case, the difficulty in awareness would be only one of many symptoms of the deficiency. We will consider the phonological alternative.

Most of the research I have mentioned thus far has concentrated on deficiencies in phonological awareness. Now, I will consider evidence that the deficiency in awareness may be symptomatic of a more general deficiency or weakness in the neurobiological device that carries out all phonological processes. In speech, phonologic structures are thus set up more weakly.

The evidence comes from comparisons of good and poor readers in their performances on tasks of short-term memory, speech perception, speech production, and naming.

Because verbal short-term memory depends on the ability to use phonological structures to hold linguistic information in memory (Conrad, 1964; Liberman, Mattingly, & Turvey, 1972), we would expect people with phonological deficiencies to have difficulties with short-term memory tasks. In many studies poor readers have, in fact, been found to have such difficulties. Typically, poor readers recall fewer items than age-matched good readers (Gathercole & Baddeley, 1988; Shankweiler, Liberman, Mark, Fowler, & Fischer, 1979; Wagner & Torgesen, 1987), but their memory difficulties occur mainly when the items

require verbal rendering. If the test materials do not lend themselves to verbal description, as in the memory for nonsense shapes or photographs of unfamiliar faces, the poor readers are not at a disadvantage (Katz, Shankweiler, & Liberman, 1981; Liberman, Mann, Shankweiler, & Werfelman, 1982).

A deficiency in the phonological processor is suggested also by the research of Brady and associates (Brady, Shankweiler, & Mann, 1983) on the speech perception of poor readers. They found that poor readers need a higher quality of signal than good readers for error-free performance in the perception of speech but not of non-speech, environmental sounds. Underlying deficits in phonological processing have also been posited by Hugh Catts (1986) to explain his finding that reading disabled students made significantly more errors than matched normals on three different tasks in which their speech production was stressed.

A similar conclusion was reached by Robert Katz (1986) in regard to the naming problems of poor readers in the second grade. The fact that poor readers tend to misname things could lead us to infer that their problem is semantic. But Katz's research with the Boston Naming Test suggests that this may be a wrong inference. The poor readers' incorrect responses to the pictured objects were sometimes nonwords closely but imperfectly resembling the target word in its phonological components ("gloav" for *glove*). Here the phonological problem is easily seen. In another kind of error, the phonological difficulty is less obvious. For example, a frequent response to the picture of a volcano was the word "tornado" which is so different in meaning that a semantic source of the error would seem likely. However, it is noted that the incorrect response has structural characteristics in common with the target word; for example, volcano has the same number of syllables, an identical stress pattern, and similar vowel constituents as tornado. More critically, it was clear that the children often actually knew the correct meaning of the word, since in subsequent questioning about the pictured object, they produced a description of a volcano, not a tornado.

Further evidence that phonological and not semantic weakness was the basis for many of the poor readers' naming errors was provided in a test of identification. Each item previously misnamed was afterwards tested for recognition by having the child select from a set of eight the one picture that best depicted the meaning of the word. In many cases, the correct object was now selected. Thus, it was possible that the poor readers had acquired internal lexical representations of many of the objects whose names they had not been able to produce accurately.

SOME IMPLICATIONS FOR INSTRUCTION

Given all that we know about the important relations between phonological ability and reading acquisition, what can we say about instructional procedures? We surely must deplore a currently popular instructional procedure, dubbed by its creators the "psycholinguistic guessing game" (Goodman, 1976). In this widely used procedure and its offshoot, the so-called "whole language" method, teachers are directed not to trouble beginners with details about how the internal phonological structure of words is represented by the letters. Instead, children are encouraged to read words however they can—for example, by recognizing their overall visual shapes—then, using their so-called normal and natural language processes, to guess the rest of the message from the context.

The "whole language" proponents seem not to have considered that before one can get to the true meaning of a sentence, one but must first get to its actual constituent words—approximations will not be enough. And to get to those actual words properly, whether one is a beginner or a skilled reader, one cannot rely on visual shape but must apply the alphabetic principle. This does not mean, incidentally, that one must necessarily sound out the words letter by letter. As we have often said elsewhere (See Liberman, Shankweiler, Liberman, Fowler, & Fischer, 1977), every reader must group the letters so as to put together just those strings of consonants and vowels that are, in the normal process of speech production, collapsed into a single pronounceable unit. There is no simple rule by which a reader can do this. Acquiring the ability to combine the letters of a new word into the appropriate pronounceable units efficiently and automatically, is an aspect of reading skill that separates the fluent reader from the beginner who has barely discovered what an alphabetic orthography is all about.

Fortunately, many children—the lucky 75 or so—do discover the alphabetic principle on their own and begin to apply it. They begin to discover for themselves the commonalities between similarly spoken and written words. When tested in kindergarten as preliterates, these children turn out to be the ones with strengths in the

phonological domain. Unfortunately, for the many children with phonological deficiencies—children who do not understand that the spoken word has segments and who have not discovered on their own that there is a correspondence between those segments and the letters of the printed word, the current vogue for the so-called (and from my point of view, misnamed) "whole language" and "language experience" approaches are likely to be disastrous. Children with deficiencies in the phonological domain who are taught in that way are likely to join the ranks of the millions of functional illiterates who stumble along, guessing at the printed message from their little store of memorized words, unable to decipher a new word they have never seen before.

For those beginners who do not discover the alphabetic principle on their own, an introductory method that provides them with direct instruction in what they need to know is critical (Liberman, 1985; Liberman & Shankweiler, 1979).

Direct instruction, as I see it, would begin with language analysis activities, which are incorporated into the daily reading lesson. These activities can take many forms, limited in number and variety only by the ingenuity of the teacher. Adaptations of three exercises that my colleagues and I advocated about ten years ago (Liberman, Shankweiler, Blachman, Camp, & Werfelman 1980) have been shown by Blachman (1987) to be successful even in inner city schools with a history of reading failure. They are outlined in Figure 1.

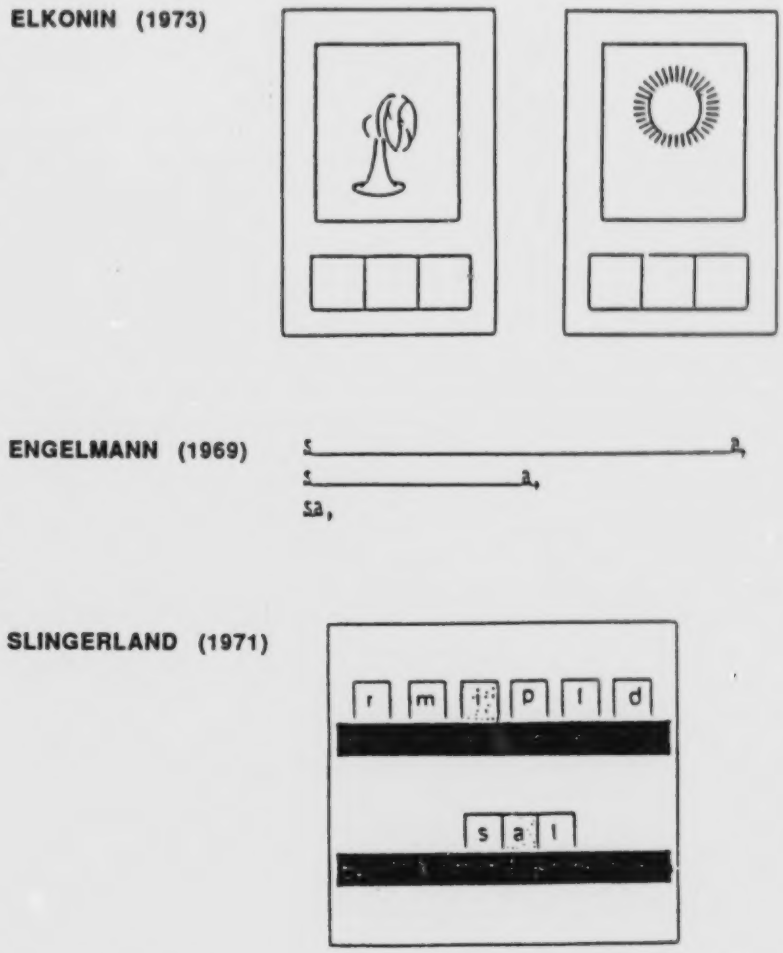


Figure 1. **Language analysis activities (after Blachman, 1987 and Liberman, Shankweiler, Blachman, Camp, & Werfelman, 1980).**

Borrowing a procedure originally devised by the Soviet psychologist, Elkonin (1973), Blachman presented the child with a simple line drawing representing the word to be analyzed. A rectangle under the drawing is divided into squares equal to the number of phonemes in the pictured word. The children are taught to say the word slowly, placing a counter in the appropriate square of the diagram as each phoneme is articulated. Later, as the child progresses, the counter can be color coded—one for vowels, another for consonants, and the letter symbols for the vowels can be added one by one. The procedure has many virtues: First, the line drawing, in effect, keeps the word in front of the children throughout the process of analysis so that they do not need to rely on short-term memory to retain the word being studied. Second, the diagram provides the children with a linear visual-spatial structure to which they can relate the auditory-temporal sequence of the spoken word, thus reinforcing the key idea of the successive segmentation of the phonemic components of the word. Third, the sections of the diagram call the children's attention to the actual number of segments in the word, so that they need not resort to uninformed guessing. Fourth, the combination of drawing, diagram, and counters provides concrete materials that help to objectify the very abstract ideas being represented. Fifth, the procedure affords the children an active part to play throughout. Finally, the color coding of the counters leads the children to appreciate the difference between vowels and consonants early in their schooling. The subsequent addition of letters to the counters can reinforce other kinds of grapheme study.

In a second activity, this one adapted from Englemann (1969), Blachman taught the children how to read the combination of a consonant followed by a vowel as a single unit. For this purpose, the initial consonant selected to be written on the blackboard by the teacher and read by the children is the continuant "s." It is chosen because, unlike the stop consonants (ptk and bdg), it can be pronounced in isolation and held over time. It is held until the teacher writes the second letter, the vowel, which is then read (as "a"). The length of time between the initial consonant and the final vowel (as well as the line drawn between them) is then reduced step by step until the two phonemes are, in effect produced as a single sound—"sa." By adding stop consonants in the final position and pronouncing the resultant words, the children can begin to collect a pool of real words (for example, *sad, sap, sag, sat*). Thereafter, new vowels and, finally, new consonants can be introduced in the same way, built into new words which subsequently can be incorporated into stories to be read and written by the child.

A similar effect can be produced by a third procedure, adapted by Blachman from Slingerland (1971). There she uses a small desk pocket-chart on which the children can manipulate individual letters to form new words and learn new phonemes. The words thus constructed, along with a few nonphonetic "sight" words, can be used in stories and poems to be read and written by the child. Note that the child is now reading and writing words the structure of which is no longer a mystery and the understanding of which can be used productively to form related words. (Note also how different this is from a common basal reader approach in which the readability level of the word *cat* is rated at Grade 1.0-2.0 but that of the word *cap* is at Grade 3.1-4.1 (Cheek, 1974) simply because *cap* happens to appear later in many basal reader series.)

All these language analysis activities and others like them can be played as games in which the introduction of each new element not only informs but delights. The beginning reader with adequate phonological ability will require only a relatively brief exposure to such activities. For such readers, these can be followed, or even accompanied by practice with interesting reading materials from other sources, and the further enhancement of vocabulary and knowledge that comes with expanded reading and life experience. But the beginners with weakness in phonological skills, as identified by the language analysis games, who may include as many as 20-25 of the children, will learn to read only if the method includes more intensive, direct, and systematic instruction about phonological structure. Research support for this view has been presented many times for at least 20 years (see Chall, 1983 or Pflaum, Walberg, Karegianes, & Rasher, 1980). It is surely time to put the research into practice.

REFERENCES

Ball, E. W., & Blachman, B. A. (1987). Phoneme Segmentation Training: Effect on Reading Readiness. Paper presented to the Annual Conference of the Orton Dyslexia Society, San Francisco.

Bertelson, P. (1987). *The onset of literacy: cognitive processes in reading acquisition.* Cambridge, MA: MIT Press.

Blachman, B. (1984). Relationship of rapid naming ability and language analysis skills to kindergarten and first-grade achievement. *Journal of Educational Psychology, 76,* 610-622.

Blachman, B. (1987). An alternative classroom reading program for learning disabled and other low-achieving children. In W.

Ellis (Ed.), *Intimacy with language: A forgotten basic in teacher education*. Baltimore: Orton Dyslexia Society.

Bradley, L., & Bryant, P. E. (1983). Categorizing sounds and learning to read—A causal connection. *Nature, 30*, 419-421.

Brady, S. A., Shankweiler, D., & Mann, V. A. (1983). Speech perception and memory coding in relation to reading ability. *Journal of Experimental Child Psychology, 35*, 345-367.

Catts, H. W. (1986). Speech production/phonological deficits in reading disordered children. *Journal of Learning Disability, 19(8)*, 504-508.

Chall, J. (1983). *Learning to read: The great debate* (updated edition). New York: McGraw Hill Book Company.

Cheek, E. H. Jr. (1974). *Cheek master word list*. Waco, Texas: Educational Achievement Corporation.

Conrad, R. (1964). Acoustic confusions in immediate memory. *British Journal of Phonology, 55*, 75-84.

Content, A., Morais, J., Alegria, J., & Bertelson, P. (1982). Accelerating the development of phonetic segmentation skills in kindergartners. *Cahiers de Psychologie, 2*, 259-269.

Cossu, G., Shankweiler, D., Liberman, I. Y., Tola, G., & Katz, L. (1988). Awareness of phonological segments and reading ability in Italian children. *Applied Psychology, 9*, 1-16.

Elkonin, D. B. (1973). U. S. S. R. In J. Downing (Ed.), *Comparative reading*. New York: MacMillan.

Engelmann, S. (1969). *Preventing failure in the primary grades*. Chicago: Science Research Associates.

Fox, B., & Routh, D. K. (1980). Phonetic analysis and severe reading disability in children. *Journal of Psycholinguistic Research, 9*, 115-119.

Gathercole, S. E., & Baddeley, A. D. (1988). The Role of Phonological Memory in Normal and Disordered Language Development. Paper presented at the Seventh International Symposium on Developmental Dyslexia and Dysphasia. Academia Rodinensis Pro Remediatone, Wenner-Gren Center, Stockholm.

Goldstein, D. M. (1976). Cognitive-linguistic functioning and learning to read in preschoolers. *Journal of Experimental Psychology, 68*, 680-688.

Goodman, K. S. (1976). Reading: A psycholinguistic guessing game. In H. Singer & R. B. Ruddell (Eds.), *Theoretical models and processes of reading*. Newark, DE: International Reading Association.

Katz, R. B. (1986). Phonological deficiencies in children with reading disability: Evidence from an object-naming task. *Cognition, 22*, 225-257.

Katz, R. R., Shankweiler, D., & Liberman, I.Y. (1981). Memory for item order and phonetic recoding in the beginning reader. *Journal of Experimental Child Psychology, 32*, 474-484.

Liberman, A. M. (1970). The grammars of speech and language. *Cognitive Psychology, 1/4*, 301-323.

Liberman, A. M. (1982). On finding that speech is special. *American Psychology, 37(1)*, 148-167.

Liberman, A. M. (1988). Reading is Hard Just Because Listening is Easy. Paper presented at the Seventh International Symposium on Developmental Dyslexia and Dysphasia. Academia Rodinensis Pro Remediatone, Wenner-Gren Center, Stockholm.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74*, 431-461.

Liberman, A. M., Mattingly, I. G., & Turvey, M. (1972). Language codes and memory codes. In A. W. Melton & E. Martin (Eds.), *Coding processes and human memory*. New York.

Liberman, I. Y. (1971). Basic research in speech and lateralization of language: Some implications for reading disability. *Bulletin of the Orton Society, 21*, 71-87.

Liberman, I. Y. (1973). Segmentation of the spoken word and reading acquisition. *Bulletin of the Orton Society, 23*, 65-77.

Liberman, I. Y. (1985). Should so-called modality preferences determine the nature of instruction for children with learning disabilities? In F. H. Duffy & N. Geschwind (Eds.), *Dyslexia: A neuroscientific approach to clinical evaluation*. Boston: Little, Brown.

Liberman, I. Y. (1987). Language and literacy: The obligation of the schools of education. In W. Ellis (Ed.), *Intimacy with language: A forgotten basic in teacher education*. Baltimore: The Orton Dyslexia Society.

Liberman, I. Y., Mann, V., Shankweiler, D., & Werfelman, M. (1982). Children's memory for recurring linguistic and nonlinguistic material in relation to reading ability. *Cortex, 18*, 367-375.

Liberman, I. Y., & Shankweiler, D. (1979). Speech, the alphabet and teaching to read. In L. B. Resnik & P. A. Weaver (Eds.), *Theory and practice of early reading*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Liberman, I. Y., & Shankweiler, D. (1985). Phonology and the problems of learning to read and write. *Remedial Special Education, 6*, 8-17.

Liberman, I. Y., Shankweiler, D., Blachman, B., Camp, L., & Werfelman, M. (1980). Steps toward literacy. In P. Levinson & C. Sloan (Eds.), *Auditory processing and language: Clinical and research perspectives*. New York: Grune and Stratton.

Liberman, I. Y., Shankweiler, D., Fischer, F. W., & Carter, B. (1974). Explicit syllable and phoneme segmentation in the young child. *Journal of Experimental Child Psychology, 18*, 201-212.

Liberman, I. Y., Shankweiler, D., Liberman, A. M., Fowler, C. A., & Fischer, F. W. (1977). In A. S. Reber & D. L. Scarborough (Eds.), *Toward a psychology of reading: The proceedings of the CUNY Conferences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Lundberg, I. (1988). Lack of Phonological Awareness—A Critical Factor in Developmental Dyslexia. Paper presented at the Seventh International Symposium on Developmental Dyslexia and Dysphasia. Academia Rodinensis Pro Remediatone, Wenner-Gren Center, Stockholm.

Lundberg, I., Frost, J., & Petersen, O.-P. (1988). Effects of an extensive program for stimulating phonological awareness in preschool children. *Reading Research Quarterly, 23/3*, 263-284.

Lundberg, I., Olofsson, A., & Wall, S. (1980). Reading and spelling skills in the first school years, predicted from phonemic awareness skills in kindergarten. *Scandinavian Journal of Psychology, 21*, 159-173.

Magnusson, E., & Naucler, K. (1987). Language disordered and normally speaking children's development of spoken and written language: Preliminary results from a longitudinal study. *Report Uppsala University Linguistics Department, 16*, 35-63.

Mann, V., & Liberman, I. Y. (1984). Phonological awareness and verbal short-term memory. *Journal of Learning Disability, 17*, 592-598.

de Manrique, A. M. B., & Gramigna, S. (1984). La segmentacion fonologica y silabica en ninos de preescolar y primer grado. *Lect. y Vida, 5*, 4-13.

Morais, J., Cluytens, M., & Alegria, J. (1984). Segmentation abilities of dyslexics and normal readers. *Perceptual Motor Skills, 58*, 221-222.

Olofsson, A., & Lundberg, I. (1983). Can phonemic awareness be trained in kindergarten? *Scandinavian Journal of Psychology, 24*, 35-44.

Olson, R., Wise, B., Conners, F., & Rack, J. (1988). Deficits in disabled readers' phonological and orthographic coding:

Etiology and remediation. Paper presented at the Seventh International Symposium on Developmental Dyslexia and Dysphasia. Academia Rodinensis Pro Remediatone, Wenner-Gren Center, Stockholm.

Pflaum, S. W., Walberg, H. J., Karegianes, M. L., & Rasher, S. P. (1980). Reading instruction: A quantitative analysis. *Educational Research, 9*, 12-18.

Shankweiler, D., Liberman,I. Y., Mark, L. S., Fowler, C. A., & Fischer, F. W. (1979). The speech code and learning to read. *Journal of Experimental Psychology: Human Learning and Memory, 5*, 531-545.

Slingerland, B. H. (1971). *A multisensory approach to language arts for specific language disability children: A guide for primary teachers.* Cambridge, MA: Educators Publishing Service.

Stanovich, K. E. (1982). Individual differences in the cognitive processes of reading: I. Word coding. *Journal of Learning Disability, 15*, 449-572.

Treiman, R. A., & Baron, J. (1981). Segmental analysis ability: Development and relation to reading ability. In G. E. MacKinnon & T. G. Walker (Eds.), *Reading research: Advances in theory and practice*, 3. New York: Academic Press.

Vellutino, F. R., & Scanlon, D. (1987). Phonological coding and phonological awareness and reading ability: Evidence from a longitudinal and experimental study. *Merrill-Palmer Quarterly, 33/3*, 321-363.

Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing in the acquisition of reading skills. *Psychological Bulletin, 101*, 192-212.

FOOTNOTE

*To appear in C. Von Euler (Ed.), *Wenner-Gren International Symposium Series: Brain and Reading*. Hampshire, England: MacMillan. (in press).

Haskins Laboratories Status Report on Speech Research
1991, SR-105/106, 9-22

The Role of Working Memory in Reading Disability*

Susan Amanda Brady†

Other presentations have focussed on the importance of phonological awareness in reading acquisition and on the central role Isabelle Liberman has played in theoretical and empirical advances in this area. It was noted that she led the way in recognizing the cognitive demands of reading: that reading, in contrast to speaking and listening, requires explicit awareness of phonological segments and that this awareness is difficult to achieve given the embedded nature of phonemes in syllables. In this paper I will show that Dr. Liberman also led the way in investigating how metaphonological abilities relate to underlying phonological processes. She was among the first to identify the need to understand the organization and functioning of the language system in order to explain sources of difficulty for poor readers. Together with Donald Shankweiler and several students, Isabelle Liberman conducted insightful and elegant research on the working memory deficits of poor readers, discovering that phonological processes are implicated here, as well as in the deficits in phonological awareness.

In my presentation I would like to accomplish three things: First, I will take note of the large body of evidence that deficits in a specifically verbal form of working memory are associated with reading problems, and to point out Dr. Liberman's enormous contribution to our understanding of the source of these limitations in language processing. Second, I will consider evidence (both old and new) that the efficiency of phonological processes is an important limiting factor in working memory capacity and that poor readers often have inefficient phonological processes. Third, I will discuss current evidence pertaining to possible causal links between the phonological processes underlying phonological awareness, verbal working memory, and lexical access.

The association of verbal working memory deficits with reading disability

While deficiency in metaphonological awareness is certainly the language factor most strongly implicated in reading disability, there is reason to believe that difficulties in awareness betoken more basic problems in language use. At the level of underlying language processes, perhaps the most striking characteristic of poor readers is the common occurrence of verbal memory problems. Teachers often comment on the difficulties poor readers have retaining information for even brief periods of time, and there now exists a massive number of research studies reporting an association between reading disability and short-term memory limitations. When given a short list of digits, or letters, or words, or even nameable pictures to recall, poor readers recall fewer items than do good readers. The generality of the relationship between reading difficulty and working memory is highlighted by the finding that it holds both for readers of alphabetic writing systems and for people in Eastern Asia learning syllabic and logographic scripts (Mann, 1985; Ren & Mattingly, 1990). Evidence also emerges from the unusual condition of hyperlexia: Healy, Aram, and Horowitz (1982) presented findings on 12 hyperlexic children who despite very low cognitive functioning in almost every area were good decoders. Performance on working memory tasks stood out as one of the few cognitive strengths of these children.

I gratefully acknowledge the helpful comments of Anne Fowler and Donald Shankweiler on an earlier draft. My research was supported in part by a Program Project Grant (HD-01994) to Haskins Laboratories from the National Institute of Child health and Human Development. Some of the material presented in this paper was adapted from the article, Reading ability and short-term memory: The role of phonological processing, by Michele Rapala and Susan Brady, 1990, in *Reading and Writing, (2)*, pp. 1-25. Dordrecht, The Netherlands: Kluwer Academic Publishers. Copyright 1990 by Kluwer Academic Publishers. Adapted by permission.

Evidence for a causal role of memory in reading performance comes from a small number of prediction studies that found memory capacity in kindergarten to be significantly related to later success at learning to read. For example, Share, Jorm, Maclean, and Matthews (1984) reported on a longitudinal study of 543 children and found a correlation of about .4 between kindergarten performance on sentence memory tasks and reading level at the end of first grade. In a subsequent analysis of these children, Jorm, Share, Maclean, and Matthews (1986) subdivided the children who became poor readers (now followed to the end of third grade) into those with a specific reading problem and those who had reading problems and also had low IQs. For both groups language deficits, including poor memory performance, were apparent at the kindergarten phase prior to reading acquisition. The low IQ children, in addition to their language impairments, had other problems such as poor performance on perceptual-motor tasks and on a nonverbal auditory task.

The consequences of a reduced capacity in working memory may be extensive. Some have suggested that limited capacity may make it more difficult to discover and master metaphonological skills. Evidence compatible with this is presented by Wagner, Balthazor, Hurley, Morgan, Rashotte, Shaner, Simmons, and Stage (1987). In this study with 111 kindergarten nonreaders, the results best fit a model specifying a single latent factor for phonological awareness and for verbal working memory. Others, such as Perfetti and Lesgold (1977; 1979), have noted that deficits in memory may make it more difficult to learn how to decode. A recent study by Dreyer, 1989, reports a correlation of .65 between performance on a recall task and decoding skill. A further impact of limited memory resources on reading has been suggested by Perfetti (1985) who hypothesized that for the slow, beginning decoder limits in working memory may be used up getting to the words of the text. Once the words have been decoded, insufficient resources are left for other higher language processes. This is consistent with the observation that children may fail to comprehend a sentence in text even when they manage to decode all the words it contains. Similarly, Liberman and her colleagues have stressed the integrative function of working memory and considered that its chief contribution would be to facilitate syntactic and other sentence-level processes (Liberman, Shankweiler, Liberman, Fowler, & Fischer, 1977; Mann, Liberman, & Shankweiler, 1980; Shankweiler, Liberman, Mark, Fowler, & Fischer, 1979). In keeping with this, Mann, Shankweiler, and Smith (1984), Smith, Macaruso, Shankweiler, and Crain (1989), and Fowler (1988) have found evidence that poor reader's difficulties with sentence level tasks stem in part from deficits in working memory.

Although memory difficulties in poor readers are commonplace, they are not universal. In an important study by Torgesen and Houck (1980), two groups of learning disabled children were identified, only one of which had memory impairments. Further, the proportion of variance between good and poor readers accounted for by working memory performance is generally markedly lower (approximately 10%) than the proportion accounted for by phonological awareness abilities (approximately 40-70%). Wagner (1988) reported on a meta-analysis of longitudinal correlational studies and training studies that finds relatively independent causal roles for metaphonological processes and for working memory. Thus the exact role of working memory deficits in reading disability remains unclear. On the one hand, the occurrence of these difficulties in poor readers is widespread, having been found across language and writing systems, in diverse populations, and to be present in those destined to be poor readers prior to reading acquisition. On the other hand, it must be acknowledged that memory limitations are not as consistently or as strongly associated with reading disability as are deficits in metaphonological skills.

These facts warrant a searching examination of working memory processes and their role in reading. Ultimately, we will need to understand the interplay between metaphonological processes and memory processes to fully comprehend the requirements of reading acquisition and the bases of reading failure.

A group of investigators from Haskins Laboratories began this enterprise by exploring the basis and extent of memory deficits for poor readers (Liberman et al., 1977; Shankweiler et al., 1979). Drawing on the evidence that information in verbal working memory is retained in a phonological or speech code, they hypothesized that poor readers have a specific difficulty with the use of phonological representations in working memory, and went on to demonstrate that good readers, like adults (Baddeley, 1966; Conrad, 1972), find lists of nonrhyming words easier to recall correctly than rhyming lists. Presumably, because in the nonrhyming case the internal phonological or speech representations for the words are very distinct, whereas for the rhyming

words they are so similar to each other that it is easier for them to be confused. Unlike the good readers, poor readers didn't show this pattern. They were not adversely affected by rhyme; performance on rhyming lists was not appreciably different from performance on nonrhyming lists. These results were interpreted to suggest that poor readers are not as adept as good readers at forming phonological representations.

In subsequent work the Haskins researchers and others have found the rhyme effect in good and poor readers to be affected by both subject and task factors (Brady, Mann, & Schmidt, 1987; Ellis, 1980; Hall, Wilson, Humphreys, Tinzmann, & Bowyer, 1983; Watkins, Watkins, & Crowder, 1974). Nonetheless, the line of research on the rhyme effect, or phonological similarity effect, paved the way in pointing to the role of phonological processes in the memory and reading difficulties of poor readers.

Using this paradigm, Shankweiler et al. (1979), made a further noteworthy observation that the memory differences between good readers and poor readers were apparent whether visual presentation was used, or auditory presentation. They suggested that poor readers have a general problem with the use of a phonological code, independent of how the material is presented, not a difficulty that is restricted to the reading process itself. In contrast, they, and others (Katz, Shankweiler, & Liberman, 1981; Liberman, Mann, Shankweiler, & Werfelman, 1982; Vellutino, Pruzek, Steger, & Meshoulam, 1973), found that poor readers do not perform less well than good readers on short-term memory tasks with nonspeech stimuli such as doodle drawings or photographs of strangers. With these stimuli, not easily given a phonological label, no significant differences in performance between good and poor readers were observed. These findings supported the conclusion that poor readers do not suffer from a general memory impairment. Rather they are often deficient in the ability to remember linguistic material, however it is presented, and the problem appears to be related to phonological processes involved in encoding or storing verbal information.

Explanations of developmental and reading-group differences in memory span

At this point it wasn't apparent whether poor readers were employing some other coding strategy to retain items in STM, such as a visual or semantic strategy, or whether they were using phonological codes, but doing so inefficiently. Long before, Conrad (1971) had suggested that children younger than six did not use phonological coding in memory. However, more recent research by Alegria and Pignot (1979) found the rhyme effect to be present in children as young as four, and, indeed, phonological coding in working memory would seem to be essential for a child to learn to speak a language. Yet, for some reason, poor readers show the rhyme effect at a later age than do children who are better readers. For example, using a cross-sectional design, Olson, Davidson, Kliegl, and Davies (1984) reported that by adolescence poor readers demonstrated the phonological similarity effect, though they still showed lower levels of recall than good readers (see also Johnston, 1982, and Siegel & Linder, 1984).

These findings suggest that factors which contribute to developmental differences in verbal STM performance may also account for reading group differences in linguistic memory. I should note that Dempster (1981), reviewing numerous studies, estimated that span approximately triples from an average of slightly more than two items at age two to an average of nearly seven items at adulthood. Children who are poor readers systematically lag behind their age mates in recall level. Three kinds of explanations have been offered to account for developmental and reading group differences in verbal STM performance. I would like to describe all three, but I'll be brief about the first two, which seem inadequate, and I'll elaborate more on the third position; an operational efficiency view that emphasizes the role of phonological processes.

Mnemonic strategies. First, the differential use of mnemonic strategies has been offered as an explanation for developmental increments in STM span. According to this view, children become increasingly aware that certain strategies will enhance recall, and are thus increasingly able to employ an appropriate technique. It is clear that mnemonic strategy use does advance as children become older. These advances include more frequent spontaneous use of rehearsal strategies, such as the way we repeat a phone number to remember it briefly, and better use of imposed or subjective organization.

Good and poor reader differences have also been explained in this way. Tarver, Hallahan, Kauffman, & Ball, 1976, and Torgesen, 1977, reported that good readers are more likely to use self-initiated rehearsal than children with reading problems. And Torgesen (1978-79) reported that good readers are more likely to use a chunking

strategy or to consciously impose an organizational plan on materials to be recalled. But additional findings, as well as interpretive difficulties, suggest that differences in mnemonic strategy use is not the sole basis for individual differences in memory function. Lange (1978) confirmed that adults and adolescents were more likely than younger children to make use of organizational principles, but differences in this strategy were not discerned in children from 5 to 12 years of age. Yet this is a period during which short-term memory span has been found to increase dramatically (Dempster, 1981).

A second conflicting piece of evidence for the role of mnemonic strategies in developmental increases in span comes from a series of experiments which have equated children's use of certain control processes during STM tasks. For example, Samuel (1978) grouped items in internal patterns in recall lists. Children ranging in age from six to nineteen all benefitted *equally* from that grouping technique, so the age differences were maintained (see also Huttenlocher & Burke, 1976, and Engle & Marshall, 1983).

Similar results have been obtained when the subjects were good readers and poor readers. That is, when reading groups are equated on the use of strategies, differences in recall span still occur. Thus while the use of mnemonic strategies improves as individuals get older, it does not appear to be the central factor either in developmental or in reading-group differences.

Capacity. A second approach to developmental changes in memory span has postulated an actual difference in capacity of short-term memory. Proponents of capacity explanations hold that a certain number of memory slots are available prior to the presentation of any stimuli. Developmental differences in span are attributed to the presence of a greater number of these slots as a function of age: so a two-year-old might have 2 slots, and adults might have 7 slots (Pascual-Leone, 1970, and Halford & Wilson, 1980). This position is consistent with the observation of developmental increases in recall, and also fits the observed differences in STM span among good and poor readers.

However, a series of experiments conducted by Baddeley, Thomson, and Buchanan (1975) demonstrated that the number of items that can be recalled is not fixed for a given individual. The number varies with the length of the stimulus (in number of syllables). Thus, temporal duration of the item list proved to be a strong determinant of span. Similarly, Ellis and Hennelley (1980) did experiments with bilingual Welsh students and showed that these subjects had a longer digit span when the digits were given in English than in Welsh, maintaining that the differences were attributable to the fact that Welsh numbers take a longer time to say. In addition, as will be discussed below, increases in memory span with age can be offset if unfamiliar stimuli are used for recall. Thus it seems than an explanation in terms of a fixed number of slots cannot account for important individual differences in STM performance.

Operational efficiency. The third position I wish to present, which I find a more promising explanation, is an operational efficiency view that proposes that memory is served by a limited capacity system and that memory operations such as encoding and retrieval become more efficient with relevant experience. As a result, the amount of operating resources needed to complete short-term memory tasks decreases with age and there is a functional increase in storage capacity. I find it useful to view the working memory system as a pie: if perception or encoding requires one quarter of the pie, three-quarters will be left over for recall. If one gets better at encoding, more resources will be available for memory. My version of this hypothesis is that the difficulty observed in encoding phonological information is not restricted to memory tasks but occurs at a more abstract level, whenever it is necessary to create and maintain a phonological representation. To evaluate this hypothesis one can look for parallels between performance on tasks involving speech perception, speech production, and verbal short-term memory. I will begin with some observations from the developmental literature and I will review here some studies with single age groups that seem relevant.

First, with adult subjects it has been observed that if the encoding requirements on a memory task are made more difficult by changing the perceptual demands, then short-term recall suffers. For example, Rabbitt (1968) had adults listening to digits in white noise. In one condition they were asked to repeat individual items, in another condition they had to recall lists of digits. Noise levels that caused no effect on perception of the individual items, significantly impaired recall of the same stimuli. Recall of the lists not in noise was greater than recall of the lists with noise. Adding noise and increasing the perceptual difficulty adversely effected memory. Likewise, Luce, Feustel and Pisoni (1983) demonstrated that adults have poorer recall of lists of synthetically spoken items than of lists of natural speech, and

Mattingly, Studdert-Kennedy, and Magen (1983) reported poorer recall if the words were spoken in different dialects. These three studies illustrate different ways of making the encoding demands more difficult, but all are compatible with the view that when more resources are used up for encoding, less are available for storage or recall, hence there is a trade off.

A second line of evidence for the relevance of underlying phonological processes to working memory functioning comes from a report by Locke and Scott (1979) that children with articulation disorders also have impaired short-term memory performance. Similarly, Ellis (1979) reviews findings of concordant error patterns by adults in spontaneous speech production and in short-term memory tasks.

The third source of evidence come from indications that for both adults and children, the intrinsic efficiency of phonological processes correlates positively with memory span. Thus significant correlations are obtained between how fast one can speak and how much is recalled on a short-term memory task. This has been found both in adult studies (Baddeley, et al., 1975; Hoosian, 1982; Naveh-Benjamin & Ayres, 1986; Nicolson, 1981) and in developmental studies. Hulme, Thomson, Muir, and Lawrence (1984) tested individuals from four years of age to adulthood and obtained a linear relationship between maximal speaking rate and recall performance. A particularly compelling set of findings comes from a study by Case, Kurland, and Goldberg in 1982. Testing three-to-six year old children, all of whom are unlikely to use mnemonic strategies, these authors observed a strong correlation between how rapidly the children could repeat the test words and the size of their memory span. The older children, who could articulate faster, recalled more of the words. In a convincing test of this relationship, Case et al. found that when adults' speaking rate was slowed to the rate for six-year-old children by giving the adults more difficult items, memory span with these stimuli also dropped to the six-year-old level. Although the encoding processes involved in memory were only evaluated by word-repetition speed, these results indicate that developmental increases in memory span may be linked to the efficiency of related phonological processes.

Let us consider the reading-group literature to examine whether the verbal memory differences of good and poor readers might also be accounted for by the efficiency of underlying phonological processes. This remains to be fully addressed: speech perception, verbal memory, and speech production abilities in good and poor readers have generally been looked at in isolation, and so relations among these processes have not been examined in depth. I'll discuss current findings addressing the operational efficiency hypothesis and will then present findings from a study with Michele Rapala (Rapala & Brady, 1990).

First, I have noted investigations which found a correspondence in poor readers between less effective use of phonological coding in short-term memory (indicated by a lack or reduction of the phonological similarity effect) and reduced memory span (Mann et al., 1980; Mark, Shankweiler, Liberman, & Fowler, 1977; Olson et al., 1984; Shankweiler, et al., 1979; Siegel & Linder, 1984). Related to this, error analyses have indicated that while both reading groups use phonological coding strategies, poor readers are less accurate. In 1983 Donald Shankweiler, Virginia Mann and I first noted this, looking at the errors made by third-grade good and poor readers on lists of non-rhyming words (Brady, Shankweiler, & Mann, 1983). The response sequences included items, for both good readers and poor readers, that had not occurred in the original strings. In the vast majority of cases, these errors were recombinations of phonological components that were present in the initial sequence (for example, for the target items *train* and *plate*, several subjects reported "trait" and "plane"). These transposition errors occurred for both reading groups, but more frequently for the poor readers. Thus, the poor readers were clearly using a phonological coding strategy, but having more difficulty than the good readers in retaining the correct combination of phonological segments. In a follow-up study, Brady et al. (1987) used lists of nonsense syllables to assess the incidence and circumstances of such errors with greater precision. We constructed lists in which the items shared zero, one, or two phonological features. Across lists of syllables, the order of the stimuli also varied systematically so we could ask whether adjacent items were more likely to be transposed than were nonadjacent ones, and whether the effect of shared features is greatest for immediately adjacent segments. As expected, poor readers made more errors than good readers, but for both groups transposition errors accounted for the majority of errors. For both reading groups, these more often involved swaps from adjacent syllables and there was a significant effect of phonological feature similarity, with a greater number of transpositions occurring between syllables that had features in common. (The effects of

feature similarity and of adjacency seemed to be independent.) These kinds of errors suggest that the inferior performance of poor readers is not the consequence of a different coding strategy, but relates to differences in the formation or storage of phonological representations. Thus, we again have a correspondence between less effective coding in STM and poorer recall.

A second area of investigation has demonstrated speech perception deficits for poor readers, possibly reflecting a general difficulty in encoding language. Different paradigms have been used such as word repetition tasks and categorical perception tasks, but the common finding is that poor readers are less accurate (Apthorp, 1988; Brady et al., 1983; Brady, Poggie, & Rapala, 1989; Godfrey, Syrdal-Lasky, Millay, & Knox, 1981; Goetzinger, Dirks, & Baer, 1960; Palley, 1986; Read, personal communication; Snowling, 1981; Snowling, Goulandris, Bowlby, & Howell, 1986; Werker & Tees, 1987.

The task has to be somewhat demanding in order to detect lower performance by poor readers. For example, Brady et al. (1983) found poor readers to be worse at identifying monosyllabic words in noise, but not monosyllabic words presented without noise. In subsequent studies we and other have also found poor readers to do less well on clearly presented words if the phonological demands are increased either by lengthening the stimuli by using multisyllabic words or by decreasing item familiarity by using pseudowords. The evidence with multisyllabic real words, as well as earlier findings that word frequency of stimuli has a comparable effect on both reading groups (Brady et al., 1983), argues against an explanation that poor readers' difficulties in phonological processing are confined to nonwords (cf., Snowling et al., 1986). It is unclear, however, whether the problem repeating words arises during the perception or production components of such tasks, or to the common requirement of formulating a phonological representation. Indeed, with respect to speech production, there are reports in the clinical literature that individuals with reading difficulty often display misarticulations in their speech (Blalock, 1982; Chasty, 1986; Johnson & Mylebust, 1967; Klein, 1986), as well as empirical demonstrations that dyslexics are slower and less accurate at repeating phrases (Catts, 1986; Catts, submitted: though for this task memory requirements may be a factor.) Finally, additional evidence of a link between phonological processes and memory capabilities with respect to reading level has been tentatively supported by positive correlations between how fast children can name lists of digits and their memory span (Spring & Capps, 1974; Spring & Perry, 1983; Torgesen & Houck, 1980.) The naming task involves speech production but also requires retrieving the names of the digits. Thus, more than one factor may be contributing to the correlation with memory.

In our own work we've found poor readers to be as fast as good readers at single word repetition, but not as accurate (Brady et al., 1989; Rapala & Brady, 1990). A recent study by Stanovich, Nathan, and Zolman (1988) using a reading-age match design makes the point that speech of articulation is strongly age-dependent. Children matched on reading ability who were in the third, fifth, and seventh grades differed markedly in speed of repetition of simple words though their memory spans were comparable. This would seem to undercut the importance of the link between reading ability and articulation, and by extension, the role of speed of formulating phonological representations in memory span. Yet the studies with adults cited earlier that report a continuing association between speed of articulating, memory span and reading ability raise doubts that developmental factors artifactually account for the correspondence between memory span and articulation found by Hulme et al. (1984). Likewise, the experimental design of Case et al. (1982) also presents convincing evidence of a significant relation between encoding processes and memory span. In sum, phonological difficulties in memory, in speech perception, and in speech production have been observed in children with reading difficulties and there have been occasional reports of correlations between these measures. Relations among the underlying phonological processes need to be more closely investigated, both in developmental studies and in reading-group comparisons. In addition, we need to disentangle the factors of speed and accuracy of articulation as they relate to memory span and reading ability.

To review, in seeking an explanation of the short-term memory deficits characteristic of poor readers, I have considered the explanations proposed for developmental changes in memory performance. The explanation in terms of mnemonic strategy use did not seem adequate. The use of mnemonic strategies does increase developmentally and is sometimes observed to be superior in good readers as compared to poor readers, but noteworthy changes in memory span occur during years in which changes in mnemonic strategies are not evident. Further, when different age

groups or reading groups are induced to use the same strategy, differences in span are still present. The capacity view, maintaining that the actual number of slots in memory increases, is contradicted by evidence that the length and familiarity of items are important determinants of recall. The third position maintains that verbal working memory is a limited capacity system and that the efficiency of encoding and retrieving limit the resources available for retaining information. Both the developmental research and the reading-group research point toward a role of the efficiency of phonological processing in memory capacity. However, in each field the questions have been incompletely addressed, and some of the findings appear to be in conflict. The developmental studies have focussed on the correspondence between speaking rate and memory, emphasizing the importance of speed of articulation. The reading studies have more extensively investigated phonological processes in speech perception and speech production that may relate to memory performance, but have generally examined each area in isolation.

A study by Michele Rapala and me (Rapala & Brady, 1990) was designed to investigate more thoroughly whether developmental and reading group differences in verbal STM can be accounted for by differences in the efficiency of related phonological processes. Using an extensive battery of tests, a cross-sectional developmental study was carried out with 4 1/2 year olds, 6 1/2 year olds and 8 1/2 year olds to examine the association between verbal STM and phonological processes in speech perception and speech production. Complementing the developmental study, a comparison of 8 1/2 year old good readers and poor readers was conducted using the same test battery to assess directly whether poor readers' difficulties in STM are associated with deficits in other phonological processes.

The measures included a verbal short-term memory task using lists of words and a non-verbal memory task known as the Corsi Blocks test (Corsi, 1972). In that task, nine identical black blocks are scattered on a platform and the experimenter points to a number of blocks, in turn. The subject then must reproduce this sequential pattern. We also included a number of phonological tasks. There were two speech repetition tasks; repetition of monosyllabic and multisyllabic words. The children were told to say each word carefully but as quickly as possible. To control for the role of perception in the first task, we also included tasks in which the child would hear a tone and was asked to respond with a particular word (in one condition the monosyllable "cat"; in another condition the multisyllable "banana"). Here no speech signal has to be encoded, so we eliminated the speech perception requirement. In the word repetition and control tasks, responses were scored for speed of onset and for accuracy. Lastly, we had a production measure in which the child was told to repeat six times in succession a two-syllable tongue twister such as "sishi" or "bublu." On this task, the time to produce the six repetitions was measured and the accuracy of each syllable was recorded.

We obtained significant correlations between all the measures in our developmental study, but of course everything improves with age. To check that the variable of age, or some general cognitive factor that improves with age, wasn't the actual basis for the significant relations, we conducted a series of correlational analyses with the effects of age controlled. When age is controlled (see Table 1) a significant relationship continues to exist between VSTM and the other phonological processes. The negative direction of the coefficients indicates that as the time to produce stimuli decreases, and as the number of errors decreases, verbal short-term memory span increases. The control measures which tested speed of articulation without a perceptual component did not correlate significantly after age was partially out. Perhaps these tasks weren't sufficiently sensitive; alternatively, the processes tapped by these measures are less closely related to memory.

If we look at the results between these measures and non-verbal memory (see Table 2), again with age partialled out, the overall patterns is very different. One variable produced a significant correlation, but with this many comparisons that could be a chance result. The lack of a general relationship between nonverbal memory and phonological processes contrasts sharply with the consistent association found between verbal recall and the other phonological measures. This indicates that while there are age-related improvements in a variety of cognitive skills, the observed relationship between verbal STM and the other phonological measures expressly reflects shared linguistic processing factors. Indeed, the results of a discriminant function analysis, conducted with the nine language measures indicated that one significant discriminant function accounted for 83.4 percent of the total variance. This result suggests the tests of speech perception, speech production, and verbal memory tap a single underlying dimension.

Table 1. *Developmental study: First order correlations (age partialled) between verbal short-term memory and each of eight phonological variables (N=74).*

Variable	Partial r
Accuracy measures	
Speech Perception: Errors on Monosyllabic Words	-.25*
Speech Perception: Errors On Multisyllabic Words	-.43*
Tongue Twister Errors	-.42*
Speed measures	
Speech Perception: RT on Monosyllabic Words	-.25*
Speech Perception: RT on Multisyllabic Words	-.32*
Tongue Twister Speed	-.35*
Control: Monosyllabic RT	-.21
Control: Multisyllabic	-.17

*p < .05

Table 2. *Developmental study: First order correlations (age partialled) between verbal short-term memory span and nine language measures (N=74).*

Variable	Partial r
Verbal STM	.20
Accuracy Measures	
Speech Perception: Errors on Monosyllabic Words	-.09
Speech Perception: Errors On Multisyllabic Words	-.12
Tongue Twister Errors	-.07
Speed measures	
Speech Perception: Reaction Time (RT) on Monosyllabic Words	-.24*
Speech Perception: RT on Multisyllabic Words	-.18
Tongue Twister RT	-.08
Control: Monosyllabic RT	-.01
Control: Multisyllabic	-.02

*p < .05

On comparing the 8 1/2 year old good and poor readers, we found, as have others, that the groups differed on verbal memory but not on nonverbal memory. We also found that memory performance correlated significantly with the error scores for multisyllabic words and for tongue twisters, and that a substantial proportion of the variance in VSTM span could be accounted for by the accuracy of phonological processing (VSTM, multisyllabic errors: $r^2 = .25$; VSTM, tongue twister errors: $r^2 = .15$).

In sum, it is likely that verbal working memory plays a major role in reading, whether memory deficits, per se, emerge as a primary factor or as a contributing factor of reading disability. Therefore, it is essential to understand the functioning of the phonological system serving verbal working memory. The present findings are com-

patible with the view that: a) poor readers are deficient in phonological processing; b) that phonological limitations lead, in turn, to limitations in the efficient use of working memory; and, c) that nonverbal memory processes are served by separate cognitive functions. However, this is a very minimal conceptual framework, and much work remains to be done on the nature of the working memory system and on the factors accounting for differences in capacity. For example, as noted earlier, the role of both speed and accuracy in phonological processing needs to be further studied to determine how these processing variables relate to the functioning of the phonological system and to the construct of efficiency. In addition, it is important to study further the basis and extent of the commonalities across speech perception, speech production and verbal memory. Our findings fit the hypothesis that the need to encode the stimulus phonologically is shared across all the tasks, and that this may be the basis of the correlations we obtained. Yet we appreciate that other factors such as rehearsal and retrieval need to be be studied in this light to evaluate alternative explanations.

In addition, the present findings suggest it would also be worthwhile to explore further whether the deficits of poor readers on verbal memory tasks are to be explained in terms of lower efficiency on a continuum of normal phonological processing. Lastly, it will be important to relate the deficits of poor readers on these more basic phonological processes to the robust evidence that poor readers also have deficits in metaphonological awareness. In the next section we turn to this question, reviewing current findings on the relations among various phonological processes.

What are the relations among phonological processes?

In other chapters in Brady and Shankweiler (in press), we have focussed on phonological awareness and on verbal working memory as potential factors in reading disability. Before we consider how these may be related, I must remind you that a third area of language function, lexical access, is often also associated with reading ability. Naming, as a test of lexical access, requires the subject to retrieve a phonological label in response to visual stimuli such as colors, numbers, letters, or pictured objects. Poor readers tend to be slower on tasks requiring the rapid naming of visual stimuli and they also have been reported to make more errors in retrieving phonologically complex labels (i.e., words such as *thermometer* or *stethoscope* (Catts, 1986; Katz, 1985). The strength of the association between reading ability and measures of lexical access has been found to depend on the particular measures taken and other factors such as word frequency, age, and task variables. The correspondence is strongest when tested in kindergarten, and when the task is to scan and label an array of high frequency items selected from a finite pool (e.g., Wolf, 1986). With older children, reading-group differences are consistently found only when the items to be labeled are orthographic symbols (e.g., Blachman, 1983; Katz & Shankweiler, 1985; Wolf et al., 1986). (See Stanovich, 1985, and Wagner & Torgesen, 1987, for reviews.)

Although each of these abilities draws on phonological representation, we are not yet in a position to explicate the nature of the underlying relations among phonological awareness, verbal working memory, and lexical access. Research addressing this question has been limited, piecemeal, and contradictory. If one considers how the phonological processes may be related, there are five logical possibilities, as Wagner and his colleagues indicate (Wagner et al., 1987): 1) There may be a unitary deficit common to metaphonological processes, phonological coding in working memory, and phonological coding in lexical access; 2) Though all are phonological, each may be a separate ability, with three independent processes implicated; 3) The metaphonological processes may be independent of the other two. Wagner et al., 1987, term this pattern of organization "an awareness versus use" classification. The remaining two combinations represent the other possible permutations of two latent abilities; 4) Phonological awareness and verbal working memory may be tied to a common factor. In this view, phonological awareness tasks rely heavily on the efficiency of coding in working memory, but use of phonological information to access stored items in the lexicon taps separate processes; 5) Phonological awareness and lexical access may be linked, perhaps by some relation between awareness of phonological structure and organization or utilization of phonological information in the lexicon. Phonological coding to maintain information in working memory would then constitute a separate aspect of processing.

The unitary deficit view is attractive in light of the compelling evidence that poor readers have deficits largely confined to the phonological domain of language. It is appealing to propose

that the three classes of phonological abilities have some underlying common factor that accounts for the individual deficits. All draw on the same underlying knowledge base: the phonological component of the language apparatus. In this way, there would be unique requirements for metaphonological, working memory, and lexical processes, but the underlying phonological representation might be inadequate and yield difficulties in each of the other areas (Liberman & Shankweiler, 1985).

So far, the studies that have attempted to examine abilities in all three phonological areas have not supported a unitary deficit, but there has not otherwise been much agreement between them. One study, a meta-analysis of the results of longitudinal correlational studies and training studies, indicated relatively independent causal roles in the acquisition of reading skills for the three phonological abilities we have identified (Wagner, 1988). The second study, conducted with 51 adults (including familial dyslexics, clinic identified dyslexics, and normally reading adults), also supported a multiple factors model (Pennington, Van Orden, Kirson, & Haith, in press), though individual tasks loaded somewhat differently on the three factors. The third, a study of preschoolers' phonological processing abilities (Wager et al., 1987) reported that at this prereading stage two coherent phonological abilities were evident: 1) memory and awareness having a common factor and 2) lexical processes as a separate component.

Other investigations have examined the relations between only two of the three phonological abilities; here too the results have been inconsistent. One line of research has attempted to examine whether the deficits in phonological awareness and memory relate to a common factor. In one view, phonological awareness tasks may rely heavily on the efficiency of coding in working memory. And indeed, those awareness measures that most strongly relate to reading ability (Stanovich, Cunningham & Cramer, 1984; Yopp, 1988) generally involve comparing a number of stimuli, manipulating segments, or deleting segments. All of these tasks require greater working memory involvement than less discriminating phonological awareness measures such as rhyme generation. Conversely, the development of phonological awareness may facilitate the ability to use phonological segments in working memory, as opposed to larger articulatory units such as syllables (see Fowler, this volume). Thus, one research goal has been to investigate whether there is a common deficit and to contrast metacognitive processes from more basic language processes. It is known, for instance, that in the absence of reading instruction, abilities to isolate and manipulate phonological segments are limited (Morais, Cary, Alegria, & Bertelson, 1979; Read, Zhang, Nie, & Ding, 1986). And yet it is presumed that the underlying phonological processes involved in ordinary speaking and listening have developed normally. Further, in studies of adults who failed to develop skill in reading, phonological awareness deficits are present, while difficulties in the other phonological abilities are not as apparent as in young poor readers (see Pennington et al., in press). These age-specific differences among poor readers lead to the hypothesis that perhaps metaphonological processes are distinct from the underlying phonological domain and that the associations between awareness and memory in young children merely reflect simultaneous increases in development. An alternative may be that the emergence of phonological awareness and subsequent reading acquisition depend on a certain minimum of working memory efficiency, beyond which individual differences in memory capacity are less crucial. In fact, research in children designed to evaluate the correspondence has yielded mixed results. Thus, Mann and Liberman (1984), Fowler (1988) and Goldstein (1976) reported significant correlations between awareness and memory tasks. On the other hand, several studies failed to find significant correlations between awareness and memory measures (Alegria, Pignot, & Morais, 1982; Blachman, 1983; Mann, 1984), or evidence (from factor analysis) that they load on a single factor (Mann & Ditunno, in press).

Nevertheless, as I've argued in the preceding section, a link appears to be justified among the more basic language tasks associated with reading disability. Verbal memory, phonological perception, speech production, and lexical access all require the creation of a phonological representation, whether this representation is initially generated internally or activated by incoming stimuli. As noted, we and others have obtained significant correlations between the accuracy of phonological processes in perception and production and the capacity of working memory; this relationship obtains developmentally as well as in comparisons of good readers and poor readers. In contrast, these processes are not related to nonverbal memory performance (cf. Rapala & Brady, 1990). Significant correlations

have also been reported between naming speed and memory span (Spring & Capps, 1974; Spring & Perry, 1983; Torgesen & Houck, 1980), although Stanovich (1985) notes inconsistencies in this outcome. Along similar lines, Elbro (1988) finds that poor performance in discriminating minimal pairs of words in noise is associated with slow lexical assess. Thus while we have evidence of correspondence between the underlying processes, these may or may not be separable from phonological awareness.

In sum, it seems premature to attempt to state how the various phonological processes are related to one another. At present, the field is characterized by lack of accord. Theoretical and empirical evidence can be gathered for a number of the potential patterns of relations among the phonological processes. Gains will no doubt be made, as they were in previous decades (e.g., Vellutino, 1979), by improving how we ask the important questions. Several methodological shortcomings are now recognized as detracting from the merit of many studies. For example, the lack of convergent results may stem from the frequent use of a single task for each construct, which may not provide an adequately valid, reliable, or sensitive measure. Thus in future work it will be important to include multiple measures of each construct to attain more robust information. In addition, it will be important to factor out general cognitive ability since this affects performance on almost all cognitive measures, and may inflate the association between any of the phonological processing tasks. Third, many of the studies have not had a sufficient number of subjects to warrant the statistical procedures used and to obtain reliable information on the questions addressed. In addition, we need conceptual advances so that we know better how to evaluate the processing requirements of various tasks. At present, in each area of phonological processing several measures are used and, for the most part, we know little about their comparability. For example, memory tasks may involve a continuous presentation task in which only the last few of a numerous and indeterminate number of stimuli are requested for recall, they may entail recall of entire strings, or of individual items. The stimuli may be digits, words, nonwords, or pictures. The task requirements are not constant, and performance on them may not vary uniformly with respect to reading ability.

Selection criteria for subjects also remains a thorny issue. We still have the concerns about how to assess reading ability and IQ, and about whether we should assess performance in other areas of achievement (e.g., math) or of cognitive functions (e.g., attention). The more recent vogue of reading-age matches only exacerbates these concerns. It has been very profitable to compare the particular reading skills of individuals who are at an equivalent reading level in order to address questions about deviant reading patterns, but I am concerned that when we extend the use of this paradigm to study underlying cognitive processes (e.g., verbal memory) that the dual diversity of age and cognitive ability makes it difficult to match subjects appropriately or to interpret the outcome (see Shankweiler, Crain, Macaruso, & Brady, in press, for further discussion).

All of these points bring me to the somewhat tiring conclusion that in the reading field we need to refine our knowledge of the individual constructs of speech perception, speech production, verbal working memory, and lexical access. In other chapeters in the Brady and Shankweiler (in press) volume, we can see that the effort to fine-tune our understanding of the construct of phonological awareness is already underway. I don't want to imply that we shouldn't at the same time attempt to test hypotheses about how the whole picture fits together, but it will be important to be cautious about interpreting individual studies. As Stanovich (1985) discusses, we need to adhere to the principle of converging evidence. The power of this principle has been awesome concerning the evidence that reading problems are associated with language deficits and, in particular, with phonological awareness. When we ask, as we have been here, how the different phonological processes relate, convergence is lacking and we need to ask why.

In closing, in this paper I have attempted to accomplish three goals. In the first section, the widespread association of verbal short-term memory deficits with reading disability was briefly reviewed. The prevalence of this association underscores the need to examine the factors that contribute to working memory performance. In the second section of the paper I argued that studies of memory development and of memory deficits in poor readers provide tantalizing indications that efficiency of phonological coding may be an important factor in memory performance. In the third section I raised the question of how working memory processes are related to other phonological processes, phonological awareness and lexical access that

have also been associated with reading deficits. At present, attempts to delineate the connections between them have yielded contradictory indications. Conceptual and methodological advances, which hinge on the continued cooperation of disciplines involved in the study of language, will be needed to advance our grasp of these issues.

REFERENCES

Alegria, J., Pignot, E. (1979). Genetic aspects of verbal mediation in memory. *Child Development, 50*, 235-238.

Alegria, J., Pignot, E., & Morais, J. (1982). Phonetic analysis of speech and memory codes in beginning readers. *Memory & Cognition, 10* 451-456.

Apthorp, H. (1988). *Phonetic coding in reading disabled community college students.* Unpublished doctoral dissertation, University of Connecticut, Storrs.

Baddeley, A. D. (1966). Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *Quarterly Journal of Experimental Psychology, 18*, 362-365.

Baddeley, A., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning & Verbal Behavior, 14*, 575-589.

Blachman, B. (1983). Are we assessing the linguistic factors critical in early reading? *Annals of Dyslexia, 33*, 91-109.

Blalock, J. (1982). Persistent auditory language deficits in adults with learning disability. *Journal of Learning Disabilities, 15*, 604-609.

Brady, S., Mann, V., & Schmidt, R. (1987). Errors in short-term memory for good and poor readers. *Memory & Cognition, 15*, 444-453.

Brady, S., Poggie, E., & Rapala, M. M. (1990). Speech repetition abilities in children who differ in reading skill. *Language and Speech, 32(2)*, 109-122.

Brady, S. A., Shankweiler, D., & Mann, V. A. (1983). Speech perception and memory coding in relation to reading ability. *Journal of Experimental Child Psychology, 35*, 345-367.

Case, R., Kurland, D. M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology, 33*, 386-404.

Catts, H. (submitted). *Phonological deficits in reading disorder children.*

Catts, H. (1986). Speech production/phonological deficits in reading-disordered children. *Journal of Learning Disabilities, 19*, 504-508.

Chasty, H. (1986). What is dyslexia? A developmental language perspective. In M. Snowling (Ed.), *Children's written language difficulties* (pp. 11-27). Windsor, Berkshire, UK: Nfer-Nelson.

Conrad, R. (1971). The chronology of the development of covert speech in children. *Developmental Learning and Verbal Behavior, 5*, 398-405.

Corsi, P. M. (1972). *Human memory and the medial temporal region of the brain.* Unpublished doctoral dissertation, McGill University.

Dempster, F. N. (1981). Memory span: Sources of individual and developmental differences. *Psychological Bulletin, 89*, 63-100.

Dreyer, L. (1989). *The relationship of children's phonological memory to decoding and reading ability.* Unpublished doctoral dissertation, Columbia University.

Elbro, C. (1988). Morphemic awareness among dyslexics. Paper presented at the Seventh International Symposium on Developmental Dyslexia and Dysphasia. Academia Rodinensis Pro Remediatone. Stockholm: Wenner-Gren Center.

Ellis, A. W. 1979). Speech production and shot-term memory. In J. Morton & J. C., Marshall (Eds.), *Psycholinguistics 2: Structure and process.* Cambridge, MA: MIT Press.

Ellis, A. W. (1980). Errors in speech and short-term memory: The efforts of phonetic similarity & syllable position. *Journal of Verbal Learning & Verbal Behavior, 19*, 624-634.

Ellis, N. C., & Hennelly, R. A. (1980). A bilingual word-length effect: Implications for intelligence testing and the relative ease of mental calculation in Welsh and English. *British Journal of Psychology 71*, 43-52.

Engle, R. W., & Marshall, K. (1983). Do developmental changes in digit span result from acquisition strategies? *Journal of Experimental Child Psychology, 3*, 429-436.

Fowler, A. (in press). How early phonological development might set the stage for phoneme awareness. In S. Brady & D. P. Shankweiler (Eds.), *Phonological processes in literacy: A tribute to Isabelle Y. Liberman.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Fowler, A. (1988). Grammaticality judgments and reading skill in grade 2. *Annals of Dyslexia, 38*, 73-84.

Godfrey, J. J., Syrdal-Lasky, A. K., Millay, K. K., & Knox, C. C. (1981). Performance of dyslexic children on speech perception tests. *Journal of Experimental Child Psychology, 32*, 401-424.

Goldstein, D. M. (1976). Cognitive-linguistic functioning and learning to read in preschoolers. *Journal of Experimental Psychology, 68*, 680-688.

Goetzinger, C., Dirks, D., & Baer, C. J. (1960). Auditory discrimination and visual perception in good and poor readers. *Annals of Otology, Rhinology, and Laryngology, 69*, 121-136.

Halford, G. S., & Wilson, W. H. (1980). A category theory approach to cognitive development. *Cognitive Psychology, 12*, 356-411.

Hall, J. W., Wilson, K. P., Humphreys, M. S., Tinzmann, M. B., & Bowyer, P. M. (1983). Phonemic-similarity effects in good vs. poor readers. *Memory & Cognition, 51*, 520-527.

Healy, J., Aram, D., Horowitz, S., & Kessler, J. (1982). A study of hyperlexia. *Brain and Language, 17*, 1-23.

Hoosian, R. (1982). Correlation between pronunciation speed and digit span size. *Perceptual and Motor Skills, 55*, 1128.

Hulme, C., Thomson, N., Muir, C., & Lawrence, A. (1984). Speech rate and the development of short-term memory span. *Journal of Experimental Child Psychology, 38*, 241-253.

Huttenlocher, R., & Burke, D. (1976). Why does memory span increase with age? *COGPSY, 8*, 1-31.

Johnston, R. (1982). Phonological coding in dyslexic readers. *British Journal of Psychology, 73*, 455-460.

Johnson, D. Myklebust, H. (1967). *Learning disabilities: Education principles and practices.* New York: Grune and Stratton.

Jorm, A. F., Share, D. L., Maclean, R. & Matthews, R. (1986). Cognitive factors at school entry predictive of specific reading retardation and general reading backwardness: A research note. *Journal of Child Psychology and Psychiatry, 27*, 45-55.

Katz, R. B. & Shankweiler, D. (1985). Phonological deficiencies in children with reading disability: Evidence from an object naming task. *Cognition, 22*, 225-257.

Katz, R., Shankweiler, D. & Liberman, I. (1981). Memory for item order and phonetic recoding in the beginning reader. *Journal of Experimental Child Psychology, 32*, 474-484.

Klein, H. (1986). The assessment and management of some persisting difficulties in learning disabled children. In M. Snowling (Ed.), *Children's written language difficulties* (pp. 59-79). Windsor, Berkshire, UK: Nfer-Nelson.

Lange, G. (1978). Organization-related processes in children's recall. In P. A. Ornstein (Ed.), *Memory development in children.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Liberman, I. Y., Mann, V. A., Shankweiler, D., & Werfelman, M. (1982). Children's memory for recurring linguistic and nonlinguistic material in relation to reading ability. *Cortex, 18,* 367-375.

Liberman, I. Y., & Shankweiler, D. (1982). Phonology and the problems of learning to read and write. *Remedial and Special Education, 6,* 8-17.

Liberman, I. Y., Shankweiler, D., Liberman, A. M., Fowler, C. A., & Fischer, F. W. (1977). Phonetic segmentation and recoding in the beginning reader. In A. S. Reber & D. L. Scarborough (Eds.), *Toward a psychology of reading: The proceedings of the CUNY Conferences* (pp. 207-225). Hillsdale, NJ: Lawrence Erlbaum Associates.

Locke, J. L., & Scott, K. K. (1979). Phonetically mediated recall in the phonetically disordered child. *Journal of Communication Disorders, 12,* 125-131.

Luce, P. A. Feustel, T. C., & Pisoni, D. B. (1983). Capacity demands in short-term memory for synthetic and natural speech. *Human Factors, 25,* 17-32.

Mann, V. A. (1984). Longitudinal prediction and prevention of reading difficulty. *Annals of Dyslexia, 34,* 117-137.

Mann, V. A. (1985). A cross-language perspective in the relation between temporary memory skills and early reading ability. *Remedial and Special Education, 6,* 37-42.

Mann, V. A. & Ditunno, P. (1990). Phonological deficiencies: Effective predictors of future reading problems. In G. Pavlikes (Ed.), *Dyslexia: A neuropsychological and learning perspective.* New York: Wiley.

Mann, V. A., & Liberman, I. Y. (1984). Phonological awareness and verbal short-term memory: Can they presage early reading problems? *Journal of Learning Disabilities, 17,* 592-599.

Mann, V. A., Liberman, I. Y., & Shankweiler, D. (1980). Children's memory for sentences and wordstrings in relation to reading ability. *Memory & Cognition, 8,* 329-335.

Mann, V. A., Shankweiler, D., & Smith, S. T. (1984). The association between comprehension of spoken sentences and early reading ability: The role of phonetic representation. *Journal of Child Language, 11,* 627-643.

Mark, L. S., Shankweiler, D., Liberman, I. Y., & Fowler, C. A. (1977). Phonetic recoding and reading difficulty in beginning readers. *Memory & Cognition, 5,* 623-629.

Mattingly, I. G., Studdert-Kennedy, M. & Magen, H. (1983, May). Paper presented at the meeting of the Acoustical Society of America, Cincinnati, Ohio.

Morais, J., Cary, L., Alegria, J., & Bertelson, P. (1979). Does awareness of speech as a sequence of phonemes arise spontaneously? *Cognition, 7,* 323-331.

Narveh-Benjamin, M. & Ayres, T. (1986). Digit span, reading rate, and linguistic relativity. *Quarterly Journal of Experimental Psychology, 38A,* 739-751.

Nicolson, R. (1981). The relationship between memory span and processing speed. In M. Friedman, J. P. Das, & N. O'Connor (Eds.), *Intelligence and learning* (pp. 179-184). Plenum Press.

Olson, R. K., Davidson, B. J., Kliegl, R., & Davies, S. E. (1984). Development of phonetic memory in disabled and normal readers, *Journal of Experimental Psychology, 37,* 187-206.

Pascual-Leone, J. (1970). A mathematical model for the transition rule in Piaget's developmental stages. *Acta Psychologica, 32,* 301-345.

Palley, S. (1986). *Speech perception in dyslexic children.* Unpublished doctoral dissertation, The City University of New York.

Pennington, B. F., Van Orden, G., Kirson, D., & Haith, M. (in press). What is the causal relation between verbal STM problems and dyslexia? In S. Brady & D. P. Shankweiler (Eds.), *Phonological processes in literacy: A tribute to Isabelle Y. Liberman.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Perfetti, C. A. (1985). *Reading ability.* New York: Oxford University Press.

Perfetti, C. A., & Lesgold, A. M. (1979). Coding and comprehension in skilled reading and implications for reading instruction. In L. B. Resnick & P. A. Weaver (Eds.), *Theory and practice of early reading.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Perfetti, C. A., & Lesgold, A. M. (1979). Discourse comprehension and sources of individual differences. Cognitive processes in comprehension. In M. A. Just & P. A. Carpenter (Eds.), *Cognitive processes in comprehension* (pp. 141-183). Hillsdale, NJ: Lawrence Erlbaum Associates.

Rabbitt, P. M. A. (1968). Channel-capacity, intelligibility and immediate memory. *Quarterly Journal of Experimental Psychology, 20,* 24-248.

Rapala, M. & Brady, S. (in press). Reading ability and short-term memory: The role of phonological processing. *Reading and Writing.*

Read, C., Zhang, Y., Nie, H., & Ding, B. (1986). The ability to manipulate speech sounds depends on knowing alphabetic transcription. *Cognition, 24,* 31-44.

Ren, N., & Mattingly, I, (1990). Short-term serial recall performance by good and poor readers of Chinese. *Haskins Laboratories Status Report on Speech Research, SR-103/104,* 153-164.

Samuel, A. G. (1978). Organization versus retrieval factors in the development of digit span. *Journal of Experimental Child Psychology, 26,* 308-319.

Shankweiler, D., Crain, S., Brady, S., & Macaruso, P. (1990). Identifying the causes of reading disability. In P. B. Gough (Eds.), *Reading acquisition.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Shankweiler, D., Liberman, I. Y., Mark, L. S., Fowler, C. A., & Fischer, F. W. (1979). The speech code and learning to read. *Journal of Experimental Psychology: Human Learning and Memory, 5,* 531-545.

Share, D., Jorm, A., Maclean, R., & Matthews, R. (1984). Sources of individual differences in reading acquisition. *Journal of Educational Psychology, 76,* 1309-1324.

Siegel, L. S., & Linder, B. A. (1984). Short term memory processes in children with reading and arithmetic learning disabilities. *Developmental Psychology, 20,* 200-207.

Smith, S. T., Macaruso, P., Shankweiler, D., & Crain, S. (1989). Syntactic comprehension in young poor readers. *Applied Psycholinguistics.*

Snowling, M. (1981). Phonemic deficits in developmental dyslexia. *Psychological Research, 43,* 219-234.

Snowling, M., Goulandris, N., Bowlby, M., & Howell, P. (1986). Segmentation and speech perception in relation to reading skill: A developmental analysis. *Journal of Experimental Child Psychology, 41,* 489-507.

Spring, C., & Capps, C. (1974). Encoding speed, rehearsal and probed recall of dyslexic boys. *Journal of Educational Psychology, 66,* 780-786.

Spring, C., & Perry, L. (1983). Naming speed and serial recall in poor and adequate readers. *Contemporary Educational Psychology, 8,* 141-145.

Stanovich, K.E. (1985). Explaining the variance in reading ability in terms of psychological processes: What have we learned? *Annals of Dyslexia, 35,* 67-96.

Stanovich, K. E., Cunningham, A. E., & Cramer, B. B. (1984). Assessing phonological awareness in kindergarten children: Issues of task comparability. *Journal of Experimental Child Psychology, 38,* 175-190.

Stanovich, K. E., Nathan, R. G., & Zolman, J. E. (1988). The developmental lag hypotheses in reading: Longitudinal and matched reading-level comparisons. *Child Development, 59,* 71-86.

Tarvar, S. G., Hallahan, D. P., Kauffman, J. M., & Ball, D. W. (1976). Verbal rehearsal and selective attention in children with learning disabilities. A developmental lag. *Journal of Experimental Child Psychology, 22,* 375-385.

Torgesen, J. K. (1977). Memorization processes in reading disabled children. *Journal of Educational Psychology, 69,* 571-578.

Torgesen, J. K. (1978-79). Performance of reading disabled children on serial memory tasks: A selective review of recent research. *Reading Research Quarterly, 14,* 57-87.

Torgesen, J. K. & Houck, D. G. (1980). Processing deficiencies of learning-disabled children who perform poorly on the digit span test. *Journal of Educational Psychology, 72,* 141-160.

Vellutino, F. R. (1979). *Dyslexia: Theory and research.* Cambridge, MA: MIT Press.

Vellutino, F. R., Pruzek, R., Steger, J., & Meshoulam, U. (1973). Immediate visual recall in poor and normal readers as a function of orthographic-linguistic familiarity. *Cortex, 8,* 106-118.

Wagner, R. K. (1988). Causal relations between the development of phonological processing abilities and the acquisition of reading skills: A meta-analysis. *Merrill-Palmer Quarterly, 34,* 261-279.

Wagner, R., Balthazor, M., Hurley, S., Morgan, S., Rashotte, C.,Shamer, R., Simmons, K. & Stage, S. (1987). The nature of prereaders' phonological processing abilities. *Cognitive Development, 2,* 355-373.

Wagner, R. K., & Torgesen, J. (1987). The nature of phonological processing in the acquisition of reading skills. *Psychological Bulletin, 101,* 192-212.

Watkins, M. J., Watkins, O. C., & Crowder, R. G. (1974). The modality effect in free and serial recall as a function of phonological similarity. *Journal of Verbal Learning and Verbal Behavior, 13,* 430-447.

Werker, J. & Tees, R. (1987). Speech perception in severely disabled and average reading children. *Canadian Journal of Psychology, 41,* 48-61.

Wolf, M. (1984). Naming, reading and the dyslexias: A longitudinal overview. *Annals of Dyslexia, 34,* 78-115.

Wolf, M., Bally, H., & Morris, R. (1986). Automaticity, retrieval processes, and reading. A longitudinal study in average and impaired readers. *Child Development, 57,* 988-1000.

Yopp, H. (1988). The validity and reliability of phonemic awareness tests. *Reading Research Quarterly, 23,* 159-177.

FOOTNOTES

*To appear in S. Brady & D. P. Shankweiler (Eds.), *Phonological processes in literacy: A tribute to Isabelle Y. Liberman.* Hillsdale, NJ: Lawrence Erlbaum Associates (in press).

†Also Department of Psychology, University of Rhode Island.

Haskins Laboratories Status Report on Speech Research
1991, SR-105/106, 23-42

Working Memory and Comprehension of Spoken Sentences: Investigations of Children with Reading Disorder*

Stephen Crain,† Donald Shankweiler,† Paul Macaruso,† and Eva Bar-Shalom†

1 INTRODUCTION

Our goal is to investigate the role of the verbal working memory system in sentence comprehension, by presenting a model of working memory in sufficient detail to allow specific predictions to be made and tested. In testing this account, we draw on experimental methods that have recently been used in research on language development. These methods are designed to control the various sources of potential difficulty in the standard laboratory tasks used to assess children's grammatical knowledge and their use of this knowledge in sentence comprehension. We illustrate how our proposals about working memory, together with the recent innovations in method, allow us to infer that abnormal limitations in phonological processing, and not absence of grammatical knowledge, is at the root of the difficulties in spoken sentence understanding that are apparent in children with reading disability.

Since reading problems are most transparent at the beginning stages of learning to read, we focus our attention there, by investigating the linguistic abilities of poor readers in the early school years. By "poor readers" we mean children who show a marked disparity between their measured level of reading skill and the level of performance that might be expected in view of their intelligence and opportunity for instruction. Our research compares performance by these children with age-matched controls—children who are proceeding at the expected rate in the acquisition of reading skills (for discussion of the issues regarding subtypes of reading disability and choice of control groups, see Shankweiler, Crain, Brady, & Macaruso, in press).

Much of the research on poor readers finds the source of their problems in the language domain, not in the area of visual perception or in general analytic ability. We shall take this for granted (for reviews, see Perfetti, 1985; Shankweiler & Liberman, 1972; Vellutino, 1979). Within the language domain, many sources of evidence converge on the conclusion that poor readers' problems reflect deficiencies in phonological processing (see Liberman & Shankweiler, 1985, and Stanovich, 1982, for reviews). However, there is one finding which raises the possibility that their limitations extend beyond phonological processing to syntactic processing as well: the discovery that poor readers characteristically fail to comprehend complex spoken sentences accurately under some circumstances. This finding has led researchers to the hypothesis that these children have not mastered all of the complex syntactic properties of the adult grammatical system (Byrne, 1981; Fletcher, Satz, & Scholes, 1981; Stein, Cairns, & Zurif, 1984). We have called this the *structural lag hypothesis* (SLH).

Portions of this research were supported by a Program Project Grant to Haskins Laboratories from the National Institute of Child Health and Human Development (HD-01994). We wish to thank the second-grade students and teachers, reading instructors, and administrators at the Coventry, CT, elementary schools. We also thank Suzanne Smith for her help with Experiment III, Henry Hamburger for extensive discussion of the issues raised in this paper, and Brian Butterworth, Myrna Schwartz and Tim Shallice for their comments on an earlier draft.

The SLH provides a coherent account of some factors that may make reading hard to learn and which may distinguish good and poor readers. This hypothesis attributes poor readers' difficulties in spoken language comprehension to their level of attainment in the acquisition of syntax. According to the SLH, language is acquired in stages, beginning with simple syntactic structures and culminating only when the most complex structures have been mastered. To explain why language acquisition conforms to a developmental schedule, the SLH endorses the idea that syntactic structures are ordered in inherent complexity. The late emergence of a structure in the course of language development is taken as an indicator of its relative complexity as compared to structures that appear earlier.

It is clear that the SLH deserves serious consideration. Reflecting some common assumptions about language acquisition and linguistic complexity, this hypothesis makes the following prediction about the language related difficulties of poor readers: the linguistic structures that beginning readers and unsuccessful older readers will find most difficult are just those that appear last in the course of language acquisition. Thus, the SLH would point to findings of language acquisition studies that suggest that some syntactic structures emerge later than others in language development, and to studies showing the late mastery of these structures by poor readers.[1] Though the SLH gives a plausible account of some of the difficulties encountered by poor readers, it has a major limitation. It gives no way to tie together poor readers' problems at the level of the sentence with their problems at the level of the word. Specifically, the postulated syntactic deficit of poor readers is independent of their deficit in processing phonological information. This means that the SLH abandons the possibility of achieving a unitary explanation of the whole symptom picture of reading disability.

In our research we have sought support for an alternative hypothesis, which we call the *processing limitation hypothesis* (PLH).[2] In contrast to the SLH, this hypothesis attempts to tie together all of the symptoms of the poor reader, viewing them as derived from inefficient processing of phonological structures. Several problems can be securely tied to a deficiency in phonological processing, including the difficulties that poor readers have in word segmentation, object naming and verbal working memory. Consider first their well-attested problems in bringing phonologic segments to consciousness. It has been shown in several language communities that on analytic tests requiring conscious manipulations of the phonemic structure of spoken words, poor readers are less proficient than children who are more successful in learning to read (Bradley & Bryant, 1983; Cossu, Shankweiler, Liberman, Tola, & Katz, 1988; Lundberg, Oloffson, & Wall, 1980; Morais, Cluytens, & Alegria, 1984). Another problem that has claimed a good deal of attention is their impaired performance on tests of object naming (Denckla & Rudel, 1976; Jansky & de Hirsch, 1972; Wolf, 1981). Analysis of the errors reveals that the mistakes are often based on phonological confusions rather than on semantic confusions (Katz, 1985). This suggests that this problem, too, is a manifestation of underlying phonological impairment.

This same line of reasoning also applies to verbal working memory. Because the verbal working memory system depends on the ability to gain access to phonological structure and use it to (briefly) maintain linguistic information, we might expect people who have phonological difficulties to show various limitations on tests of ordered recall (Baddeley, 1986; Conrad, 1964, 1972; Liberman, Mattingly, & Turvey, 1972). For poor readers, as in other language-impaired populations, there is ample evidence in the literature testifying to deficiencies in short-term retention of verbal materials. Differences in recall have been obtained with a variety of verbal materials, including words and spoken sentences, but they are not typically found with materials that cannot be coded linguistically (see Liberman, Shankweiler, Liberman, Fowler, & Fischer, 1977; Wagner & Torgesen, 1987). Moreover, there is direct evidence from memory experiments that poor readers in the beginning grades are less affected by phonetic similarity (rhyme) than age-matched good readers. This is another indication of their failure to fully exploit phonologic structure in working memory (Mann, Liberman, & Shankweiler, 1980; Olson, Davidson, Kliegl, & Davies, 1984; Shankweiler, Liberman, Mark, Fowler, & Fischer, 1979).

In addition to these symptoms, we noted earlier that poor readers are sometimes unable to comprehend spoken sentences as well as comparable good readers. The central aim of this paper is to explain how the difficulties of poor readers in understanding spoken sentences may be derived from deficient phonologic processing. On the face of it, these difficulties might seem to require another kind of of explanation. But suffice it to say here that the findings of our recent

research, including the results of the experiments presented in Section 4, have persuaded us that the source of their spoken language comprehension failures is also tied to an underlying deficiency in phonological processing, as proposed by the PLH, and is not the result of a lag in syntactic development, as predicted by the SLH.

Given these sharply contrasting hypotheses about poor readers' problems in sentence comprehension, we now turn to the kinds of evidence that can decide between them. One source of evidence may be obtained by examining the pattern of errors good and poor readers make in response to sentences of different types. If poor readers suffer from a limitation in processing, it makes sense that the pattern of errors on different structures should be similar for both groups, with the poor readers showing a decrement of roughly the same magnitude on each sentence type. The prediction that the error pattern of poor readers should parallel that of good readers serves as the foundation for one of the experiments reported in Section 3.

Another research strategy which has proven useful in distinguishing between the PLH and SLH is to examine the performance of good and poor readers on laboratory tasks which differ in how severely they tax the resources of working memory. Marked improvement in performance in the face of reduction in memory load is anticipated by the PLH but not by the SLH. In the absence of requisite structures, poor readers should fail in comprehension even when memory load is minimal. On the other hand, if a processing limitation is the source of the problem, even the most unskilled reader should prove competent with highly complex linguistic constructions in spoken language, within the constraints imposed by their limitations in processing capacity. This prediction too is tested in the experiments we report below. Before we give details of the experiments, it will be useful to describe our view of the working memory system and its role in language processing.

2 ORGANIZATION OF THE LANGUAGE APPARATUS

Our conception of the language apparatus shares much common ground with the modularity proposal advanced by Fodor (1983). It grows out of a biological perspective on language that has long guided research on speech at Haskins Laboratories. According to this viewpoint, the language faculty functions autonomously in the sense that it is supported by special brain structures and operates according to principles that are specific to it and not shared by other cognitive systems. One source of evidence for this conception of modularity comes from studies of speech perception (Mattingly & Liberman, 1988). Another source is from the study of aphasia and related disorders where there is evidence that a circumscribed lesion in the left hemisphere may selectively perturb certain aspects of language performance, leaving other linguistic and nonlinguistic abilities relatively intact (Linebarger, Schwartz, & Saffran, 1983; Marin, Saffran, & Schwartz, 1976; Shankweiler, Crain, Gorrell, & Tuller, 1989). There is also evidence that ability to process language may be preserved in the face of massive losses to other systems, as in cases of "isolation aphasia" (e.g., Whitaker, 1976).

Another source of evidence for modularity comes from the study of language development, where it has been found that complex linguistic principles emerge in young children at a characteristic pace that is independent of the emergence of other cognitive systems or principles (e.g., Hamburger & Crain, 1984). Also important are research findings demonstrating children's early mastery of linguistic principles that go beyond the data provided by the environment (e.g., Crain & McKee, 1985; Crain & Nakayama, 1987; Crain, Thornton, & Murasugi, 1987). Taken together, all of these findings sustain the notion that language is a biologically coherent system, as the modularity proposal maintains.

An extension of the modularity proposal supposes that the language faculty itself is composed of several autonomous subcomponents (or submodules). This componential view of sentence production and comprehension postulates several structures and processors. Roughly, each *structure* is a stored system of rules and principles corresponding to a level of linguistic representation: phonology, syntax and semantics. In addition to the independent levels of structural representation, the language apparatus contains special *processors*, including the phonological, syntactic and semantic parsers. Each parser is a special-purpose device for rule access and ambiguity resolution corresponding to a specific level of representation. Each parser operates on principles and rules in assigning constituent structure to linguistic input. Because the parsers operate on constituent structure, and not on sequences of words themselves, we can understand sentences of great length, but can retain only relatively short lists of unrelated material.

Two further architectural features of the language apparatus are essential to our explanation of the difficulties poor readers have in sentence understanding. We assume, first, that the various submodules are arranged in a hierarchical fashion, with a unidirectional and vertical ("bottom up") flow of information such that a lower level passes results to higher levels but not the reverse. It is also critical to our view that transactions between the parsers take place "on-line," with the results of low level analyses being quickly discarded, to make room for subsequent input (for related discussion, see Carpenter & Just, 1988).

How working memory functions in the language processing system. In keeping with the modularity hypothesis, we conceive of verbal working memory as a domain-specific system that subserves the language apparatus.[3] The primary function of verbal working memory is to facilitate the extraction of a meaning representation corresponding to the linguistic input. Assuming that the extended modularity hypothesis is correct, this involves the interaction of several structures and processors. As we conceive of it, verbal working memory is an active processing system in which the analysis of verbal material by these structures and processors takes place during language processing.

In common with other contemporary approaches, we assume that there are two components to the working memory system (Baddeley, 1986; Baddeley & Hitch, 1974; Carpenter & Just, 1988; Daneman & Carpenter, 1980; Perfetti & Lesgold, 1977). First, there is a storage buffer where rehearsal and initial (phonological) analysis of phonetically coded information takes place. This buffer has the properties commonly attributed to short-term memory. It can hold information only briefly, perhaps only for 1-2 sec, in the order of arrival, unless the material is maintained by continuous rehearsal. The limits on capacity of the buffer mean that information must be rapidly encoded in a more durable form if it is to be retained for subsequent higher level analysis. Our conception of the storage buffer bears obvious similarities to other discussions in the literature.

What is new in our conception of the verbal working memory system concerns its other component. We view this as a control mechanism whose primary task is to relay the results of lower-level analyses of linguistic input upward through the language apparatus. Its regulatory duties begin at the lowest level by bringing phonetic (or orthographic) input into contact with phonological rules, for word level analysis. Phonologically analyzed information must be rapidly transferred out of the storage buffer and shunted to the syntactic processor, at the same time freeing the storage area to accept the next chunk of phonetic material. By synchronizing information flow with input, the control mechanism is able to push results upward through the system rapidly enough to promote on-line extraction of meaning (Crain & Steedman, 1985; Marslen-Wilson & Tyler, 1980; Wingfield & Butterworth, 1984).

In processing spoken language, on-line parsing explains how individuals with drastically-curtailed working memory capacity—capable of holding only two or three items of unstructured material—are sometimes able to comprehend sentences of considerable length and complexity (Martin, 1985; 1990; Saffran, 1985). Previous research has found it paradoxical that aphasic patients with a severely restricted phonological short-term store are sometimes capable of understanding at a level far exceeding what would be expected on the basis of their span limitations. This result is fully consistent with our model of working memory.

In reading, on-line processing of syntactic and semantic representations necessarily depends on prior orthographic and phonological processing. Until the reader is proficient in decoding from print, we would expect that reading is more demanding than speech of working memory resources. Sometimes it is assumed that print confers an advantage because the reader can look back. It is important to appreciate, however, that only the skilled reader can exploit the opportunity to reexamine sentences in text which were not successfully parsed on first reading. In the unskilled reader, the working memory system is usually preoccupied with orthographic decoding.

3 IDENTIFYING THE SOURCE OF READING DISABILITY

We are now in a position to show how the architectural arrangement of the language faculty can be exploited to provide an explanation of the sentence comprehension difficulties of poor readers. A modular view of the language apparatus raises the possibility that a single component may be the source of the entire symptom complex that characterizes reading disability. It is clear that failures in sentence comprehension could arise, in principle, from a deficit (or deficits) at any level that ultimately feeds into the semantic component. It is also conceivable, however, that the entire symp-

tom complex of poor readers, including their difficulties in spoken language comprehension, implicates the phonological component. Let us explain how.

Recall that the submodules of the language faculty act in strict sequence ("bottom up") to assign a partial structural analysis, which can then be passed on to higher levels. To keep information flowing smoothly, the control mechanism must avoid unnecessary computation that would delay the rapid extraction of meaning. This means that, in ordinary circumstances, the working memory buffer need not store many segments of unanalyzed linguistic material. But suppose that the phonological analysis of material in the buffer is impeded for some reason. Given the architectural features of the language apparatus we have proposed, this would also have the effect of curtailing the operation of higher level analyses of verbal material. In short, the functions of an otherwise intact system would be depressed.

This is exactly what happens in cases of reading disability, in our view. Since poor readers are deficient in setting up and organizing phonological structures, sentence comprehension is compromised because inefficient phonological analysis creates a "bottleneck" that constricts information flow to higher levels of language processing. Although the remaining components of the language apparatus may be completely intact, their operation will be hobbled by poor readers' limitations in phonological processing. In effect, a lower-level deficit in phonologic processing masquerades as a deficit at higher levels. At this point, however, we cannot rule out the possibility that the comprehension problems of poor readers are caused by a deficiency in some other component of the language apparatus (e.g., in syntactic parsing). But since poor readers' comprehension problems follow automatically from their well-attested limitations in phonological processing, it becomes unnecessary to postulate additional impairments within the language system. Moreover, we will provide evidence of the acquisition of complex syntax for both good and poor readers, as anticipated by the modularity hypothesis (see also Shankweiler & Crain, 1986).

It is important to underscore another expectation of our model, that poor readers should display successful comprehension on sentences that are not especially taxing of phonological resources. This distinguishes our view from other proposals about the relation between working memory and sentence comprehension (e.g., Baddeley, Vallar, & Wilson, 1987; Vallar, Basso, & Bottini, 1990). As long as the control mechanism of working memory is intact, even persons with abnormal limitations in phonological short-term storage capacity should be able to understand sentences of considerable complexity, if they do not impose excessive demands on phonological resources. Since the control mechanism of working memory plays such a prominent role in explaining why impaired comprehension should appear on specific sentences and not others, it will be worthwhile to describe it in more detail.

The compiling analogy. Pursuing an analogy with the compiling of programming languages, we view the control mechanism of working memory as a *control structure* whose function is to carry out a series of translations, each being a translation from a relatively high level language (the source language) to a more detailed language (the target or object language). This concept is familiar in computer science, where high level languages like Pascal or Lisp are compiled into lower-level languages such as assembly language or machine language. But the notion of compiling is quite general, and has proven useful in modeling human language processing as well.

Cognitive compiling occurs in natural language processing experiments in which a subject is asked to act out the interpretation of a sentence using toys and figures provided in the experimental workspace. Here, the source language, e.g., English, must be translated into a more detailed language that underlies the overt actions the subject makes in response to the input. We will refer to the mental language that serves as the target language for observable physical actions as the *language of plans*. In our view, several interesting properties of the control component of working memory can be illuminated by considering the translation between input sentences and the plans that they evoke (see Hamburger & Crain, 1984, 1987, for further discussion and for empirical data).[4] In the paragraphs that follow, we focus on the difficulties that may arise for the executive component of working memory in the process of translating from language input to plans. We consider first situations that are amenable to simple translation between source and target language (Hamburger and Crain, 1984). Then we will look at particular linguistic forms which deviate from the best-case scenario, thereby exacting a toll from the resources of working memory.

In the simplest case, (i) each well-formed fragment of target language code is associated with a

single constituent of source language code, (ii) the fragments of target language code can be concatenated to form the correct representation of the input, (iii) the fragments can be combined in the same order they are accessed, and (iv) each fragment is processed immediately after it is formed, permitting the source code to be discarded. These conditions form a straightforward process of sequential look-up-and-concatenation. Rarely, however, are all the conditions met in ordinary language. And when they are not, the computations involved in reaching the target code (e.g., the semantic interpretation or plan associated with a linguistic expression) could stretch the resources of verbal working memory. It will not be possible to spell out each condition in detail, but it may be helpful to make a few remarks about each, focussing on the linguistic constructions that appear in the experiments reported in Section 4.

(i) The first condition is an isomorphism between any two levels of representation. Correspondence of this kind is maintained between syntactic and semantic constituency in Montague Grammar in order to provide a systematic account of the assignment of semantic values to linguistic expressions. The foundation of this account is the principle of *compositionality*, which states that the meaning of a linguistic expression is determined from the meanings of its constituent expressions (and their mode of combination). Despite the appeal of a straightforward relationship between the syntax and semantics, linguists working in the generative framework have argued that syntax and semantics are largely autonomous. In our terms, this is liable to add to the complexity of translating between syntactic and semantic structural representations.[5]

(ii) Whether or not the first condition is met, it seems reasonable to suppose that the simplest way to combine nodes of the target language is by concatenation. Unfortunately, it is clearly not possible to concatenate meanings even in parsing simple natural language phrases like "expensive socks" or "second bear." Since expensive socks are not expensive, it would be a mistake to evaluate this phrase on a word-by-word basis, e.g., by forming a semantic value for "expensive" (say, the set of expensive things), and then combining this with the semantic value of the following word, "socks." Similarly, the second bear is not necessarily in second position in an ordered array. On occasion, concatenation of word meanings is possible, for example with NPs that contain *absolute adjectives*, like "green," "fuzzy," "Albanian," and so on, where the denotation of the adjective is not dependent on the linguistic context (e.g., *naked Albanian wrestler*). But since no unique semantic value can be given to *relative adjectives*, (e.g., expensive) or to *ordinals*, the human sentence processing mechanism must hold off interpreting these prenominal modifiers until the head noun has been received.

Translations that require the parser to splice together dissociated pieces of code at some level also violate the simple process of look-up-and-concatenation (see Hamburger & Crain, 1984, 1987). An example of this source of distress for working memory is "second striped ball." An analysis of the logical structure of the plan corresponding to this phrase shows it to consist of a nested loop structure in which fragments of plan associated with "striped ball" are inserted into the piece of code associated with the ordinal "second." Breaking apart the code needed to increment a counter is required in order to test objects (for stripedness and ballhood), to ensure that the counter is advanced only as stripped balls are located. This process is referred to as *compiling discontinuity* by Hamburger and Crain.[6]

Empirical support for the claim that phrases like this present difficulties for young children comes from several acquisition studies which find that children often choose object (b) from an array like the following in response to a request such as "point to the second striped ball" (Matthei, 1981; Roeper, 1972). That is, children incorrectly select the object that is second and striped and a ball, instead of the second of the striped balls (d).

(a) (b) (c) (d) (e) (f)

Hamburger and Crain (1984) suggest that this error may be the spurious result of premature interpretation of "the second..." as applying to the entire set of objects in the array. They found that children were dissuaded from this *concatenative* response if they were asked to handle the subsets of objects before these were placed in the array. This presumably inhibits premature execution, since it is unclear in this circumstance which (sub)set of objects the ordinal "second" modifies.

(iii) There is another locus of difficulty in translating from a source language form to target language code: condition (iii). This condition requires the order of concatenation of plans to mirror the linguistic input. Let us call any violation of this condition a *sequencing problem*. In addition to

compiling discontinuity, a sequencing problem arises in the example of "second striped ball." As we saw, the locus of the difficulty with this phrase is not with either the source code or the object code, but only in their relationship. This suggests a possible alternative to the merging of code in the formation of a plan. The alternative would be to hold onto the code associated with the ordinal "second" until after the remaining elements have been combined (establishing the data structure for "striped ball."). But setting aside a constituent to await the preparation of other elements with which it is to associate is assumed to be costly of memory resources. In terms of the model of working memory we are considering, this would constitute a violation of condition (iii) and, as a consequence, also condition (iv).[7]

Relative clauses. A second example of the difficulties a sequencing problem may pose for comprehension is from a study on the acquisition of restrictive relative clauses. This study (Hamburger & Crain, 1982) discovered that many children who performed the correct actions associated with sentences like (1) often failed, nevertheless, to act out these events in the same way as adults.

(1) The cat scratched the dog that jumped through the hoop.

Most 3-year-olds and many 4-year-olds acted out this sentence by making the cat scratch the dog first, and then making the dog jump through the hoop. Older children and normal adults act out these events in the opposite order, the relative clause *before* the main clause. Intuitively, acting out the second mentioned clause first seems conceptually more correct since "the dog that jumped through the hoop" is what the cat scratched.

It is reasonable to suppose that this kind of conflict between the order of mention and conceptual order (and most appropriate order of execution) stresses working memory because both clauses must be available long enough to enable the hearer to formulate the plan which represents the conceptual order. Presumably, this difference between the responses of children and adults reflects the more severe limitations in children's working memory in coping with sentences that pose sequencing problems.

The response we have characterized as conceptually correct (with the relative clause acted out first) requires the formation of a two-slot template, and a specification of the particular sequence in which the actions are to be carried out. Since on the simple look-up-and-concatenate scenario, processing occurs on line (i.e., in a left-to-right word-by-word basis), it seems to us that the difficulty presented by the conflict between order-of-mention and conceptual order occurs because the information in both clauses must be held in memory long enough to put the first-mentioned action into the second slot. If memory is overloaded, a subject may adopt a default procedure of acting out clauses in their order of mention—that is, according to the simple translation routine of look-up-and-concatenate.

To explain this phenomenon we draw upon another analogy to translation among programming languages. Here we appeal to the distinction between *compiling*, which completes the translation before starting to execute, and *interpreting*, which interleaves translation and execution. We can use this distinction in explaining children's conceptually incorrect responses to sentences like (1). Since children are unable to hold information long enough in working memory to *compile* a conceptually correct plan, it makes sense to suppose that they opt instead to *interpret* in cases like (1). Consistent with this supposition is the observation that children often begin to act while the sentence is still being uttered.

Temporal terms. A third example of the sequencing problem arises with sentences containing the temporal terms *before* and *after*. These terms explicitly dictate the conceptual order of events, and they too may present a sequencing problem by introducing conflicts between conceptual order and order-of-mention. This is illustrated by sentence (2).

(2) Jabba flew the X-Wing fighter after Han Solo sped away in the Millennium Falcon.

A sequencing problem arises in (2) because the order in which events are mentioned is opposite to the conceptual order. Again, research in language acquisition has found that young children frequently interpret these sentences in an order-of-mention fashion (Clark, 1970; Johnson, 1975). As with relative clause sentences, it is likely that this response reflects an inability to hold both clauses in memory long enough to formulate a plan for acting them out in the correct conceptual order. Once again, children's failure to segregate translation and execution explains their incorrect, default decision to adopt the simple look-up-and-concatenate translation.

In this case, however, an alternative account of children's difficulties has been proposed. It has been argued that a structural explanation, and

not a processing explanation, is called for. Proponents of the structurally-based explanation (Amidon & Carey, 1972) point out that the same children who failed to act out sentences like (2) correctly emitted a high rate of correct responses to sentences similar in meaning, but with simpler syntactic structure, as in (3).

(3) Push the motorcycle last; push the helicopter first.

There is direct evidence that processing factors, and not lack of syntactic competence, are responsible for children's errors in comprehending sentences with temporal terms. The evidence is this: once processing demands are reduced, most 4- and 5-year-old children usually give the correct response to test sentences like (4) and (5).

(4) Push the helicopter after you push the motorcycle.
(5) Before you push the motorcycle, push the helicopter.

To minimize processing load, one must take cognizance of a presupposition on the use of temporal terms. The presupposition associated with sentences (4) and (5) is that the hearer intends to push a motorcycle. These sentences are felicitous only if the subject has already indicated an intent to play with a motorcycle prior to receiving the test sentence. To satisfy this presupposition, one simply has to ask the child *in advance* to select one of the toys to play with before each trial. When young children were given this contextual support, they displayed unprecedented success in comprehending sentences with the temporal terms "before" and "after" (Crain, 1982; Gorrell, Crain, & Fodor, 1986).[8] The same finding was also obtained in a recent study of mentally retarded adults (W. Crain, 1986).

We should also mention a superficial linguistic property that forestalls premature execution, and thereby eases the burden on working memory, in sentences like (5). This is the presence of a temporal term in the initial clause, which indicates that a two-slot template is required. Notice that in the corresponding sentence with "after," the temporal conjunction appears in the second clause. The account of memory difficulties we have proposed would therefore lead us to expect this type of sentence to be harder, especially if it contains "after." This prediction is confirmed in the experiments reported in section 4 below.

In our discussion of the control component of working memory, we have assigned to it as few combinatorial duties as possible. This makes it essentially a simple-minded traffic controller for symbolic representations that are being composed within the submodules of the language apparatus. It is also apparent that the structure building operations that take place within these modules are frequently at odds with the efficient management of information flow.[9] Much is gained, however, by having them incorporated into the language faculty, since they supply the generative capacity for producing and understanding (an uncountable number of) novel sentences.

Garden path effects. Corresponding to each intermediate level of representation is a processing mechanism, or parser. The task of each parser is to assign structure to the incoming code as it is being transmitted from the next-lower parser. This analysis phase of the compiling process was aptly referred to by Miller (1956) as *chunking*.

The syntactic parsing mechanism is probably the best understood of the parsers. This mechanism consists of a number of routines for accessing syntactic rules and principles and resolving ambiguities that arise when more than one analysis is compatible with the current input We assume that access to rules during on-line processing uses hard-wired portions of the language apparatus—almost reflex like in character—that are sparing of processing capacity in most cases. However, natural languages permit massive local ambiguity and, despite the flexibility that this allows, it surely incurs some cost to memory resources.

In fact, there is considerable evidence that local ambiguities are quickly resolved, perhaps within one or two words after they arise. One parsing tendency that seems to have evolved to meet the twin exigencies of ambiguity and working memory limitations is called *Right Association* (see Frazier, 1978; Kimball, 1973.) Right Association explains why listeners or readers connect an incoming phrase as low as possible in the phrase marker that has been assigned to the preceding material. This 'strategy' reflects the functional architecture of the language apparatus, which has many computations to perform and little space for their compilation and execution. As a result, strategies like Right Association dictate that incoming material is integrated into the most readily available (i.e., local) node in the phrase marker under construction. So, for example, Right Association dictates that the adverb "yesterday" will be attached to the lower of the two VPs in the ambiguous sentence (6) and will therefore be interpreted as related to the last mentioned event.

(6) Bush said he apologized to the UAW, yesterday.

In keeping with Right Association, there is a strong tendency for people to interpret (6) to mean that Bush apologized yesterday, and not that he uttered a sentence to that effect yesterday.

It is reasonable to suppose that memory limitations promote rapid on-line integration of material into a structural representation. Although parsing strategies may enable the parser to circumvent the limitations of working memory, they sometimes introduce problems of their own, because the decision dictated by a strategy may turn out to be incorrect in the light of subsequent input. In this case, the perceiver is led down a garden path by the parser. The existence of 'garden path' effects (illustrated in (7)) shows that for some sentences even full knowledge of the grammar is not powerful enough to overcome the liability of a tightly constrained working memory.

(7) Bush said that he will apologize to the UAW, yesterday.

Recovery from garden paths is possible only within the limits of working memory, because this determines whether the grammatically correct attachment site is still available. Since sentences that tax working memory heavily have been found to present problems for poor readers, they should be less able than good readers to recover from incorrect analyses prompted by parsing strategies like Right Association. Therefore, they should be even more susceptible than good readers to garden path effects. Experiment III (reported below) tests this prediction by asking good and poor readers to respond to several types of garden path sentences.

An examination of how the test sentences were constructed may help to clarify the logic of this experiment. Suppose you are looking at a picture in which a girl (Mary) is using a crayon to draw a picture of a monkey who is drinking milk through a straw. The corresponding sentence is given in (8). What is the unspecified NP in this situation? Both "a crayon" and "a straw" are grammatically well-formed, but the analysis favored by Right Association has "*with NP*" modifying "*drinking milk*" rather than modifying "*drawing a picture*," so the general preference is to cash out the NP as "a straw."

(8) Mary is drawing a picture of a monkey that is drinking milk with NP

This parsing preference is still present if the NP in (8) is extracted by Wh- Movement, as in (9).

(9) What is Mary drawing a picture of a monkey that is drinking milk with?

The preposition "*with*" again coheres strongly with the relative clause, rather than with the main clause. The result is that one is tempted to make an ungrammatical analysis of (9) in which "*what*" has been extracted from the relative clause, violating a putative universal constraint on extraction called *Subjacency*. Research in language acquisition, using a picture verification task, found that many children succumb to this temptation, in an *apparent* violation of Subjacency, responding to (9) by saying "a straw," rather than giving the grammatically correct response, "a crayon" (Crain & Fodor, 1985; Otsu, 1981).

This incorrect response clearly bears on the choice of the two hypotheses we are considering about the source(s) of reading disability. Since Subjacency is part of Universal Grammar, the PLH would maintain that it should be adhered to by good and poor readers alike from the earliest stages of language development. On the other hand, the processing limitations of poor readers would lead us to expect them to make more apparent violations of the Subjacency constraint. It is then incumbent on the PLH to show that the relatively poor performance of poor readers on sentences like (9) is due to parsing pressures (viz. the effect of Right Association) rather than to ignorance of universal constraints on syntax.

There are two critical ingredients in determining whether responses which violate the Subjacency constraint reflect a processing limitation or, instead, arise from a structural deficit. As noted earlier, if poor readers suffer from a processing limitation, this should be revealed in the pattern of errors across sentence types for both reader groups: poor readers should show a decrement in performance across sentence types, but there should be no group-by-sentence-type interaction. This pattern emerges from comparison of the responses of the reader groups in Experiment III. The final ingredient is a demonstration of the grammatical competence of poor readers with the construction under investigation. This is the objective of Experiment IV.

4 APPLYING THE WORKING MEMORY MODEL TO IDENTIFY THE CAUSES OF SENTENCE COMPREHENSION FAILURES IN POOR READERS

In this section we elaborate on the specific problems that should be incurred by poor readers, given our model of language processing. Four experiments are reported here. These experiments were designed to test between the two competing

hypotheses (sketched in Section 1) about the source of impaired comprehension of spoken sentences by poor readers. Specifically, we ask whether the sentence processing difficulties are due to a syntactic deficit, as claimed by the SLH, or alternatively, whether they reflect a limitation in processing involving working memory, as claimed by the PLH. To explore both possibilities, we selected good and poor readers in the second grade. Reader groups were established on the basis of combined word and non-word scores on the Decoding Skills Test (DST) of Richardson and DiBenedetto (1986). To ensure that the difficulties experienced by the poor reader group could not be attributed to a general deficiency in cognitive function, the reader groups were equated on intelligence as well as on chronological age. (For discussion of the general efficacy of this experimental design, see Shankweiler, Crain, Brady, & Macaruso, in press.)

Temporal terms (Experiments I and II)

In the first two experiments, we were interested to discover how variations in processing load affect the performance of poor readers relative to good readers. In the last section, we saw that sentences which contain temporal terms are of particular interest in deciding between the competing hypotheses because (i) temporal terms have been found to emerge late in the course of normal language development, and (ii) the source of late mastery has been attributed to syntactic complexity, as the SLH would suggest, as well as to their demands on memory resources, as the PLH would have it. In order to test between these hypotheses, Experiments I and II used a figure manipulation paradigm, with input sentences containing adverbial clauses with the temporal terms "before" and "after." This task engages children in a game in which they are asked to move toys as dictated by orally presented sentences. The set of objects available in the experimental workspace was the same in both experiments; it comprised nine objects (cars, trucks, horses) of different colors and sizes.

The purpose of the first experiment was to establish a baseline of linguistic competence by good and poor readers with sentences containing temporal terms. In the second experiment, we sought to manipulate processing demands in two ways. First, an additional modifier was added to one of the noun phrases in half of the test sentences. This maneuver increased the possibility that subjects would make errors in selecting the objects to be moved on each trial. A second change involved presenting the test sentences in contexts that satisfied the presupposition associated with the use of the temporal term. We hypothesized that poor readers would show appreciable performance gains when processing demands were minimized through the satisfaction of this presupposition. It should be kept in mind that if the poor reader group displayed a sufficiently high level of correct performance in any condition, this would argue against the hypothesis that the relevant syntactic structures are missing from their grammars. But, in addition, an increase in successful comprehension in felicitous contexts would lend credence to a processing explanation of their performance failures in less than optimal contexts.

Each experiment was carried out with a different set of 14 good and 14 poor readers. The mean combined reading scores (on the DST) for the good and poor readers were 92.9 and 23.7 out of 120, respectively (Experiment I), and 97.2 and 37.9 (Experiment II). The IQ of subjects was calculated on the basis of their performance on the Peabody Picture Vocabulary Test—Revised (Dunn & Dunn, 1981). Performance on this test was used to ensure that both groups were in a similar IQ range, and that the differences between good and poor readers could not be attributed to different levels of vocabulary knowledge. The mean Peabody score for good and poor readers was 110.6 and 105.0, respectively (Experiment I), and 115.4 and 109.0 (Experiment II).

Experiment I

The purpose of Experiment I was to assess the level of linguistic competence for both reader groups with sentences containing temporal terms. This experiment employed simple NPs and, like many previous studies in the acquisition literature, provided no contextual support.[10] In half of the twelve test sentences the order-of-mention of events corresponded to the conceptual order of events, as in (10). In the other half, the order in which events were mentioned was opposite to the conceptual order, as in (11).[11]

(10) Push the red car before you push the largest horse.

(11) Push the smallest horse after you push the blue car.

First of all, we found that poor readers made significantly more errors than good readers, $F(1,27) = 4.92$, $p < .04$. However, the overall performance of both groups was high, with the poor reader group performing well above chance (87.5% correct). This indicates that the poor readers were not lacking the necessary competence to successfully interpret temporal term sentences even

when they contain inessential prenominal modifiers. The near-ceiling performance of the good reader group (96% correct) meant that subsequent analyses of their error patterns would not be revealing, so the remainder of our analyses focuses on the pattern of errors by the poor readers.

In particular, we were interested in determining whether or not the sentences we expected to be most demanding of memory resources do indeed cause special problems for poor readers. These sentences are the ones which present a conflict between the conceptual order and the order-of-mention and contain the temporal term *after*, as in (11). Poor readers' 21.4% errors on these sentences reflects their highest error percentage for any sentence type. In fact, it is a significantly higher error rate than for *before* sentences (7.1% errors) of the same type, $F(1,13) = 4.50$, $p=.05$. This confirms our expectation that sentences like (11) would be the most difficult for poor readers given their inherent memory limitations.

In Experiment II, we asked whether a high proportion of correct responses is still characteristic of poor readers in contexts that are even more demanding of working memory resources. If not, the combined data would lend support to the hypothesis that poor readers suffer from a limitation in processing. This difference across tasks would defy explanation on the hypothesis that they suffer from a developmental lag in the acquisition of complex syntax.

Experiment II

The purpose of this experiment was to test the effects of varying memory demands on good and poor readers.[12] According to the account of the working memory system presented earlier, poor readers should be highly sensitive to alterations in processing load which give rise to problems in cognitive compiling. We sought first to exacerbate the processing load beyond the level imposed in Experiment I by including an additional prenominal modifier in half of the test sentences. As exemplified in (12) and (13), these sentences contained the ordinal term 'second,' which introduces discontinuity in related statements in the plan that one must compile in order to respond accurately to the noun phrase in which the ordinal appears. We will refer to sentences with NPs of this sort as *complex NPs*.

(12) Push the second smallest horse before you push the blue car.

(13) Pick up the second largest truck after you pick up the blue horse.

A second change in design was introduced in order to increase the ease of processing. We took advantage of a pragmatic property that is often associated with sentences containing subordinate clauses, namely, their presuppositional content, and we exploited this property to reduce the burden imposed on memory by satisfying the presupposition associated with test sentences of this type, as discussed earlier. In the revised procedure, children are asked, before each test sentence is presented, to identify one object they want to play with in the next part of the game. The experimenter subsequently incorporates this information into the subordinate clause introduced by the temporal term. For instance, sentence (12) would have been presented only after a subject had selected the blue car. This will be referred to as the *felicity* condition. In the other, *no felicity* condition, the presupposition inherent in the use of temporal terms was not satisfied; sentences were presented in the "null context," as in Experiment I. In the null context, unmet presuppositions must be "accommodated" into the listener's mental model of the discourse setting (Lewis, 1979). In order to compensate for unmet presuppositions, the subject must revise his/her current mental model by averring that the presupposition was met. Updating one's knowledge representation in this way is known to be costly of processing resources (see Crain & Steedman, 1985, and references therein). In light of these considerations, the PLH anticipates a high rate of successful comprehension by both reader groups in the *felicity* condition, but it predicts that poor readers' performance will suffer in contexts that are more taxing of working memory, as in the *no felicity* condition.

The stimuli in Experiment II consisted of 16 sentences with temporal terms *before* and *after*. In contrast to Experiment I, only four sentences were presented in which the order-of-mention of events was the same as their conceptual order, as in (12). In the remaining twelve sentences, order of mention was opposite to the conceptual order, as in (13). All children encountered the test sentences in both contexts, i.e., in the *felicity* and *no felicity* conditions. This required two testing sessions for each child, with half of the children receiving contextual support in the first session, and half in the second session.

Overall analyses of the results reveal main effects of *reader group* ($F(1,26) = 14.16$, $p<.001$), *felicity* ($F(1,26) = 6.50$, $p<.02$), and *NP complexity* ($F(1,26) = 6.13$, $p<.02$). In addition, there is a marginally significant NP *complexity* × *reader*

group interaction ($F(1,26) = 3.92$, $p=.06$) and a trend toward a *felicity* × *reader group* interaction ($F(1,26) = 2.89$, $p=.10$). The main effect of *reader group* tells us that poor readers performed less well than good readers. However, the main effect of *felicity* indicates that satisfying the felicity conditions, i.e., reducing the processing demands created by conflicts in sequencing, produced a significant reduction in errors for both groups. The marginally significant *felicity* × *reader* group interaction suggests that the satisfaction of presuppositions increased performance for poor readers to a greater extent than for good readers. As displayed in Figure 1, there is a greater disparity between their performance for *no felicity* than for *felicity*. This lends credence to the hypothesis that, without contextual support, poor readers' limitations in working memory are exacerbated. The fact that poor readers perform at a success rate of 82.4% when the felicity conditions were satisfied, even when half of the test sentences contained complex NPs, calls into question the claim of the SLH that poor readers lag in their mastery of complex syntactic structures.

Averaged over the *felicity* and *no felicity* conditions, the main effect of *NP complexity* tells us that complex NPs evoking significantly more errors than simple NPs. However, the marginally significant *NP complexity* × *reader group* interaction (see Figure 2) indicates that poor readers were more adversely affected by changes in NP complexity than good readers. The special difficulties that the poor readers displayed with the sentences containing complex NPs presumably reflect the fact that these sentences are more taxing on working memory resources.

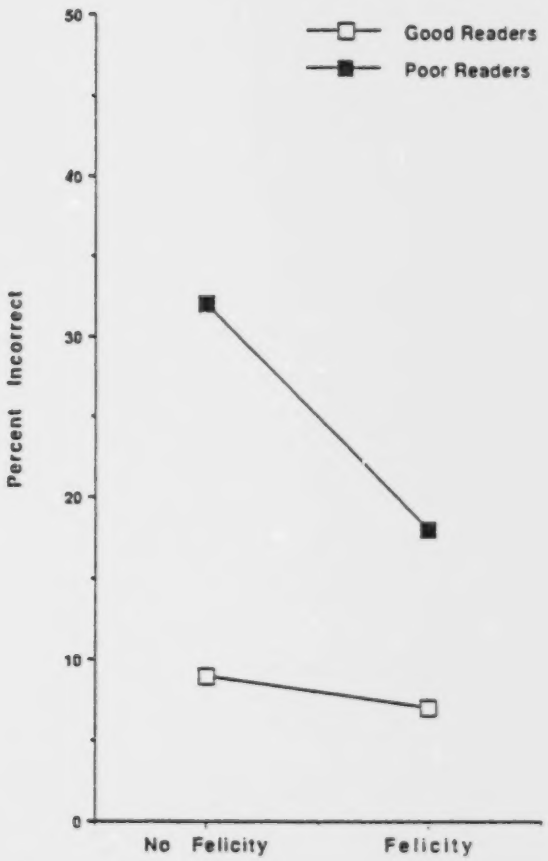


Figure 1. Percentage of incorrect responses to temporal term sentences (Experiment II).

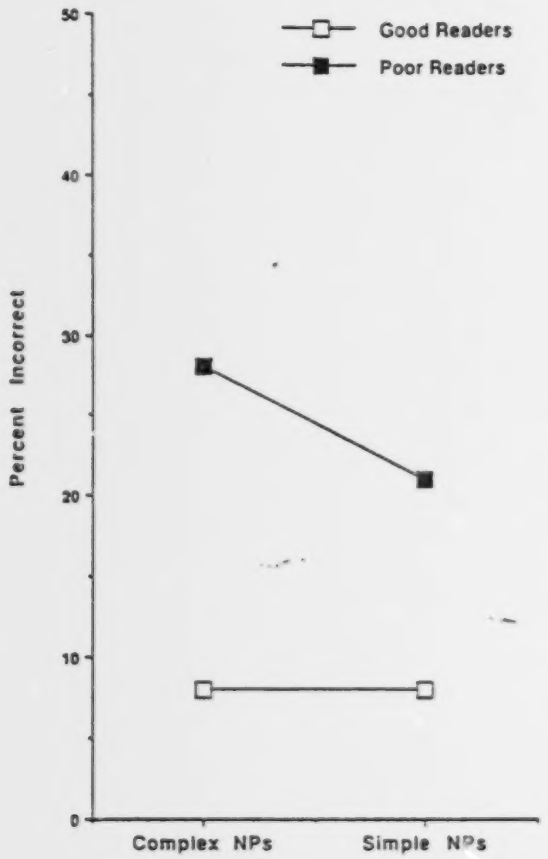


Figure 2. Percentage of incorrect responses to simple and complex NP sentences (Experiment II).

This is explained by the model of working memory presented earlier as the outcome of the cumbersome problem of compiling a plan that violates another of the conditions for simple translation from input sentence to target plan, the problem of compiling discontinuity.

As in Experiment I, we took a closer look at the sentences that we hypothesized would be the most difficult for poor readers, i.e., *after* sentences that pose a conflict between order-of-mention and conceptual order, as in (13). Restricting our analyses to the *no felicity* condition only, we find that poor readers made the most errors on just these sentences. In fact, they produced significantly more errors on them than they produced on *before* sentences of the same type, $F(1,26) = 6.86$, $p<.02$. This difference is not reflected in the good readers' errors for these same sentences under the same conditions. The significant *after vs. before* × *reader group* interaction ($F(1,26) = 5.56$, $p<.03$), as shown in Figure 3, reveals this discrepancy.

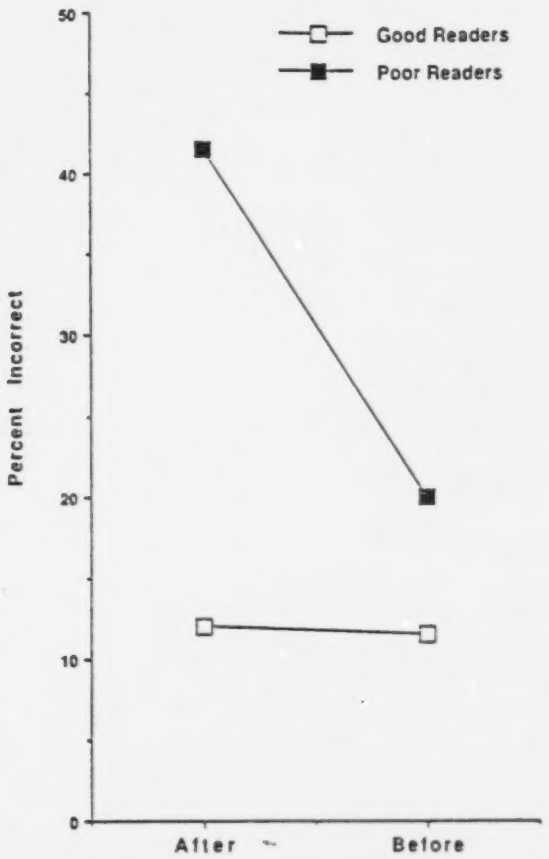


Figure 3. Percentage of incorrect responses when conceptual order conflicts with order-of-mention in the No Felicity condition (Experiment II).

As predicted, the worst case for poor readers was the *after* sentences that pose a conflict between order-of-mention and conceptual order and no contextual support in the form of satisfied presuppositions. Good readers, on the other hand, were not handicapped by the memory demands this situation places on the subject.

A glance at the type of errors made by poor readers in the most difficult situation indicates almost as many responses in which they identified the wrong object (19%) as responses in which they acted out the clauses in the wrong order (25%). In a few instances both error types occurred in processing the same sentence. However, in the corresponding situation in Experiment I (i.e., in response to *after* sentences when order-of-mention conflicts with conceptual order), poor readers made relatively few wrong object responses (7%) compared to wrong order responses (14%). Most likely, this difference in the relative frequency of error types is due to the manipulation of NP complexity across the experiments. Poor readers were more likely to choose a wrong object in sentences containing complex NPs in Experiment II, since choosing an object denoted by a complex NP requires overcoming the additional problems of compiling discontinuity.

Taken together, the findings of Experiments I and II indicate that as processing demands are increased, poor readers' performance on an object manipulation task involving temporal terms sentences is eroded much more than good readers' performance. Decreasing the processing demands, either by satisfying the felicity conditions or by using less complex NPs, elevates performance by poor readers, such that group differences diminish, and all subjects perform at a high level of success. Results showing improved performance by poor readers when processing demands are lowered cannot be explained on the hypothesis that poor readers are lacking the relevant syntactic structures. Only an account in terms of processing limitations can explain these findings. We hypothesize, based on the results of Experiments I and II, that if we conducted an additional experiment which presented only simple NPs (as in Experiment I), but with the felicity conditions satisfied, we would find near-perfect performance by both groups.

Garden path effects (Experiments III and IV).

In Experiment III we examined the responses of good and poor readers to the kind of garden path sentences discussed in section 2. As we noted, an incorrect response to these sentences can be explained in two ways, corresponding to the two

hypotheses we have been considering about the source of poor readers' problems in sentence comprehension. First, errors could reflect the absence in a subject's grammar of a structural constraint on extraction, Subjacency. Alternatively, they could be the result of an inability to overcome the effect of the parsing strategy Right Association, presumably due to limited working memory resources.

Three types of garden-path sentences were presented, in order to vary the processing load placed on working memory. As indicated earlier, varying sentence types puts us in a position to examine the error pattern across sentence types to help distinguish between the PLH and the SLH. The three types of syntactic constructions are illustrated in (14)-(16). There were four sentences of each type (adapted from Crain & Fodor, 1985).

(14) Relative Clause: Who is Bill pushing the cat that is singing to?

(15) Prepositional Phrase (deep): What is Jennifer drawing a picture of a boy with?

(16) Prepositional Phrase (distant): Who is Susan handing over the big heart-shaped card to?

Sentence (15) is labeled *deep* because the origin of the 'extracted' Wh-phrase is a prepositional phrase that is embedded in an NP which is itself embedded in an NP. This contrasts with the *distant* case (16), in which there is only one level of embedding. Although the sentences are matched for length, we anticipated that *distant* PP sentences would be easier to process than either the *deep* PP sentences or the relative clause sentences. The depth of syntactic embedding in both relative clause and *deep* PP sentences means that they deviate more than the *distant* PP sentences from the simple look-up-and-concatenation translation process presented in the last section.

A set of 44 good readers and 46 poor readers (which includes all of subjects of Experiment I and II) participated in this study. The mean combined word and non-word reading scores on the DST for the good and poor readers were 96.3 and 37.5, respectively. On each trial subjects were asked to listen carefully to a tape-recorded set of sentences which described a scene depicted in a large cartoon drawing placed in front of them. Immediately following the description, they were asked to respond to a question about some aspect of the drawing. As an example, the context sentences in (17) preceded the test question (14).

(17) Bill's father is waiting for Bill to bring him the cat. The cat loves to sing and has made up a song for his toy mouse.

The grammatically correct response to this question is "his father." The response "the mouse" is incorrect, since it represents an apparent violation of Subjacency. The PLH would argue that an examination of the pattern of errors across sentence types for each group may provide evidence that these errors are not in fact violations of Subjacency at all, but are the result of the processing burdens these sentence impose on working memory. It is difficult to say exactly what the SLH would predict about the pattern of responses by good and poor readers for any of the sentence types presented in this experiment. It seems reasonable to suppose, however, that the SLH might anticipate that the reader groups would display different profiles, since these sentences are exceedingly complex. To reiterate, the PLH predicts that both groups will manifest a similar pattern of errors across sentences of varying syntactic types, with poor readers penalized to a greater degree than good readers on sentences that are costly of working memory resources (e.g., the PP 'deep' and relative clause sentences).

This is exactly what was found. Analysis of the percentage of incorrect responses made by both groups reveals main effects of *sentence type* ($F(2,176) = 21.53$, $p<.001$) and *reader group* ($F(1,88) = 8.95$, $p<.004$), but no interaction of sentence type and reader group. Figure 4 shows that the effect of sentence type is due mainly to the higher percentage of errors for relative clause (30.5% errors) and PP deep (32.4% errors) sentences than for PP distant (15.8% errors) sentences, as anticipated by the model of working memory we presented. The reader group difference reflects the fact that good readers (21.4% errors) performed significantly better than poor readers (31.0% errors).

The absence of interaction means that poor readers show a general decrement in performance, but responded in the same way as good readers to the three different types of garden path sentences. As may be seen in Figure 4, there are no constructions which provide disproportionately greater difficulty for poor readers. This invites the inference that errors for both reader groups should receive the same interpretation. Since good readers exhibit a high proportion of correct responses on these sentences, it is reasonable to conclude that their errors are ones of performance and do not reflect an underlying deficiency in

syntactic knowledge. The significant reader group difference across sentence types might, at first glance, suggest that poor readers are lacking this knowledge, were it not for the absence of a group by sentence type interaction.[13]

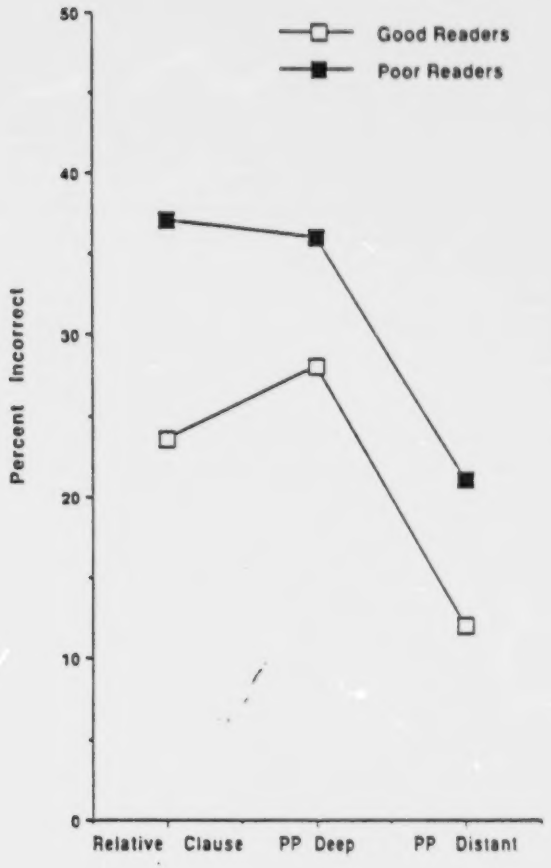


Figure 4. **Percentage of incorrect responses to garden path sentences (Experiment III).**

To support our contention that poor readers do not suffer from a lack of grammatical competence, we conducted a further experiment to assess competence with the structure which most clearly differentiated the two reader groups, the relative clause (good readers: 23.9% errors; poor readers: 37.0% errors). If the PLH is correct, and poor readers are not missing some relative clause structures, then the differences between reader groups should evaporate when competence is assessed in a way that reduces the processing burdens of working memory. This would reinforce the inference that the reader group differences in evidence here should be attributed to a processing limitation on the part of the poor reader group and not a structural deficit.

Experiment IV

Following the proposal presented in the previous discussion, this experiment used the relative clause as a further testing ground between the hypotheses. To this end, we sought to establish that the reader group differences in response to relative clause sentences in Experiment III do not reflect lack of grammatical competence on the part of poor readers, by showing that poor readers are able to successfully comprehend relative clauses in contexts that ease memory load. Ideally, we would like to provide evidence that the apparent violations of Subjacency by poor readers disappear when the demands on memory are reduced. However, reducing processing load for sentences which could induce Subjacency violations is nearly impossible. Therefore we chose an alternative approach: to assess the ability of poor readers to comprehend the most complex substructure contained in the sentences which evoked the greatest number of errors, i.e., the ones containing relative clauses. Thus, this experiment was designed to assess the competence of poor readers in interpreting relative clauses with memory demands held to a minimum. Positive results in this case would lend further support in favor of the joint claim that poor readers possess the universal constraint of Subjacency but failed to perform as well as the good readers in Experiment III because of their deficient processing capabilities.[14]

The relative clause has been a focus of research in normal language acquisition as well as in the literature on reading disability. Relative clauses are found to evoke difficulties in interpretation for preschool children (Tavakolian, 1981), and for older children who are poor readers (Byrne, 1981; Stein et al., 1984). Early research in both areas led some researchers to the conclusion that children's poor performance was due to a lack of syntactic knowledge. However, Mann et al. (1984) tested good and poor readers' comprehension of relative clauses using an act-out task and found that, although good readers performed significantly better than poor readers overall, both groups were affected in the same way by the difficulty of the type of relative clause. This familiar error pattern of good and poor readers is further evidence that they differ in processing capabilities, rather than in structural competence, as we have seen.

Further support for the view that poor readers' difficulties with relative clauses reflect performance factors comes from an additional study of

good and poor readers' comprehension of relative clauses by Smith, Macaruso, Shankweiler, and Crain (1989). Adapting several experimental innovations from the literature on language acquisition, Smith found that poor readers made few errors when the pragmatic presuppositions on the use of relative clauses were satisfied (Hamburger & Crain, 1982). Smith et al. compared their findings with those of the Mann et al. study, in which the same subject selection criteria were used, but in which the presuppositions of relative clauses were unsatisfied. Taken together, the data from these studies revealed that the changes in methodology had the effect of eliminating reader group differences, with the result that both good and poor readers performed at a high level of success.

The present study employed an act-out task which incorporated some of the methodological innovations used by Smith et al. and by Hamburger and Crain, in order to assess the grammatical competence of a subset of the good and poor readers who participated in Experiment III. The same set of 14 good and 14 poor readers from Experiment I participated in the present experiment. Three types of relative clause constructions were used.[15] Examples are provided in (18)-(20).

(18) SO The lion that the bear bit jumped over the fence.
(19) OO The boy touched the girl who the ice cream fell on.
(20) OS The lady hugged the man who picked up the suitcase.

In order to reduce the processing burden imposed by sentences such as (18)-(20), we satisfied one of the presuppositions of relative clauses. Specifically, we incorporated two objects in the experimental workspace corresponding to the head noun of the relative clause. For example, in (19), there were two figurines from which the subject could choose to act out the sentence. By including the extra figurine, we satisfied the requirement of a restrictive relative clause to restrict, in the example, the set of girls to the one on which the ice cream fell.

As in the Smith et al. study, we found no significant group differences and a low error rate for all subjects (good readers: 8.7% errors; poor readers: 11.1% errors). In addition, the pattern of errors across sentence types was virtually identical for both groups. This is shown in Figure 5. The main finding was that poor readers acted out relative clause sentences with an 89% success rate when the appropriate presuppositions were met. In fact, for two of the sentence types, poor readers made only 7% errors. This provides support for the contention that a syntactic deficit was not the cause of the inferior performance of poor readers in Experiment III. The results of Experiment IV strongly suggest that poor readers are able to comprehend various relative clause constructions. These results invite the inference that the poor readers' higher error rate in the context provided in the earlier study by Mann et al. was a consequence of their abnormal limitations in working memory.

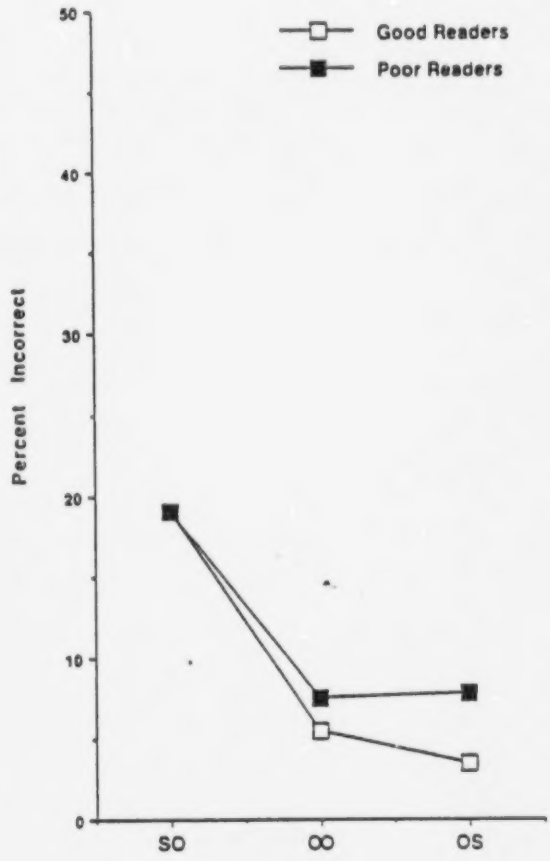


Figure 5. Percentage of incorrect responses to relative clause sentences (Experiment IV).

5 CONCLUSION

The purpose of the present research was to develop an approach to language comprehension problems that is sufficiently detailed to allow specific predictions and powerful enough to identify causes. The first half of the paper lays the groundwork. It presents the major assumptions about the language apparatus that underlie our research. Working within the framework of a modular view of language, we gave an account of verbal working memory that emphasizes its control functions more than its storage functions. We proposed that the task of the control component of working memory, in extracting meaning from linguistic forms, is to regulate the flow of linguistic material through the system of parsers, from lower- to higher-levels of representation. The inherent limitation in buffer capacity entails that higher-level processing must be executed within very short stretches of text or discourse. To deal with this problem, the control component of working memory must rapidly transfer partial results of linguistic analyses between levels of structural representation, i.e., the phonology, the lexicon, the syntax, and the semantics.

A further goal was to show how this model could be successfully applied to get to the root of difficulties that children who are poor readers often display in processing spoken sentences. Alternative hypotheses about the causes of poor readers' sentence comprehension problems were posed, and the conceptual machinery for testing between these hypotheses was introduced. The hypotheses make different predictions because they locate the source of reading difficulties in different components of the language apparatus. Roughly, the views turn on the distinction between structure and process. On the SLH, poor readers suffer from a structural deficit, i.e., a deficit in the stored mental representation of principles of syntax, in addition to their (unrelated) deficiencies in phonological processing, verbal working memory, and so on. On the PLH, each of the deficits of poor readers is a reflection of their limitations in processing phonological information.

To test between these hypotheses, we reviewed the results of four interlocking experiments. A pattern of findings emerged that indicates that the necessary syntactic structures were in place, and that the source of the poor readers' difficulties in comprehension of spoken sentences stemmed from inefficiencies in on-line processing of sentences that for one reason or another stressed working memory. Thus, inefficiency of verbal working memory and not failure to acquire critical language structures was identified as the factor responsible for the comprehension difficulties. We argued that, ultimately, the problems of poor readers originate at the phonological level, and that difficulties that might appear to reflect a syntactic deficiency are, in reality, manifestations of a special limitation in accessing and processing phonological structures.

REFERENCES

Amidon, A., & Carey, P. (1972). Why five-year-olds cannot understand *before* and *after*. *Journal of Verbal Learning and Verbal Behavior, 11*, 417-423.

Baddeley, A. D. (1986). *Working memory*. Oxford: Oxford University Press.

Baddeley, A. D., & Hitch, G. B. (1974). Working memory. In G. B. Bower (Ed.), *The psychology of learning and motivation (Vol. 8)*. New York: Academic Press.

Baddeley, A. D., Vallar, G., & Wilson, B. (1987). Comprehension and the articulatory loop: Some neuropsychological evidence. In M. Coltheart (Ed.), *Attention and performance XII*. London: Lawrence Erlbaum Associates.

Bradley, L., & Bryant, P. (1983). Difficulties in auditory organization as a possible cause of reading backwardness. *Nature, 271*, 746-747.

Brown, R. (1973). *A first language*. Cambridge, MA: Harvard University Press.

Byrne, B. (1981). Deficient syntactic control in poor readers: Is a weak phonetic memory code responsible? *Applied Psycholinguistics, 2*, 201-212.

Carpenter, P. A., & Just, M. A. (1988). The role of working memory in language comprehension. In D. Klahr & K. Kotovsky (Eds.), *Complex information processing: The Impact of Herbert A. Simon*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Chomsky, C. (1969). *The acquisition of syntax in children from 5 to 10*. Cambridge, MA: MIT Press.

Clark, E. V. (1970). How young children describe events in time. In G. B. Flores d'Arçais and W. J. M. Levelt (Eds.), *Advances in psycholinguistics*. Amsterdam: North Holland.

Conrad, R. (1964). Acoustic confusions in immediate memory. *British Journal of Psychology, 3*, 75-84.

Conrad, R. (1972). Speech and reading. In J. F. Kavanaugh & I. G. Mattingly (Eds.), *Language by ear and by eye: The relationships between speech and reading*. Cambridge, MA: MIT Press.

Cossu, G., Shankweiler, D., Liberman, I. Y., Tola, G., & Katz, L. (1988). Awareness of phonological segments and reading ability in Italian children. *Applied Psycholinguistics, 9*, 1-16.

Crain, S. (1982). Temporal terms: Mastery by age five. *Papers and Reports on Child Development, 21*, 33-38 (Stanford University).

Crain, S., & Fodor, J. D. (1985). On the innateness of Subjacency. *Proceedings of the Eastern States Conference on Linguistics, Volume 1*, The Ohio State University, Columbus, Ohio.

Crain, S., & McKee, C. (1985). Acquisition of structural restrictions on anaphora. *Proceedings of the Northeastern Linguistics Society, 16*, University of Massachusetts, Amherst, MA.

Crain, S., & Nakayama, M. (1987). Structure dependence in grammar formation. *Language, 63*, 522-543.

Crain, S., & Shankweiler, D. (1987). Reading acquisition and language acquisition. In A. Davidson, G. Green & G. Herman

(Eds.), *Critical approaches to readability: Theoretical bases of linguistic complexity*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Crain, S., & Steedman, M. (1985). On not being led up the garden path: The use of context by the psychological syntax processor. In D. R. Dowty, L. Karttunen & A. Zwicky (Eds.), *Natural language parsing: Psychological, computational, and theoretical perspectives*. Cambridge, MA: University Press.

Crain, S., Thornton, R., & Murasugi, K. (1987). Capturing the evasive passive. Paper presented at the Boston University Conference on Language Development, Boston, MA.

Crain, W. (1986). *Restrictions on the comprehension of syntax by mentally retarded adults*. Unpublished doctoral dissertation; Claremont Graduate School.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*, 450-466.

Denckla, M. B., & Rudel, R. G. (1976). Naming of object-drawings by dyslexic and other learning disabled children. *Brain and Language, 3*, 1-15.

de Villiers, J. G., Tager-Flusberg, H. B. T. Hakuta, K., & Cohen, M. (1979). Children's comprehension of relative clauses. *Journal of Psycholinguistic Research, 8*, 499-518.

Dunn, L., & Dunn, L. (1981). Peabody Picture Vocabulary Test—Revised. Circle Pines, MI: American Guidance Service.

Fletcher, J. M., Satz, P., & Scholes, R. (1981). Developmental changes in the linguistic performance correlates of reading achievement. *Brain and Language, 13*, 78-90.

Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.

Frazier, L. (1978). *On comprehending sentences: Syntactic parsing strategies*. Unpublished doctoral dissertation, University of Connecticut.

Goldsmith, S. (1980). *The psycholinguistic bases of reading disability: A study in sentence comprehension*. Unpublished doctoral dissertation, The City University of New York.

Gorrell, P., Crain, S., & Fodor, J. D. (1986). Contextual information and temporal terms. Paper presented at the Boston University Conference on Language Development, Boston, MA.

Hamburger, H., & Crain, S. (1982). Relative acquisition. In S. Kuczaj, II (Ed.), *Language development, Volume 1: Syntax and semantics*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Hamburger, H., & Crain, S. (1984). Acquisition of cognitive compiling. *Cognition, 17*, 85-136.

Hamburger, H., & Crain, S. (1987). Plans and semantics in human processing of language. *Cognitive Science, 11*, 101-136.

Jansky, J., & de Hirsch, K. (1972). *Preventing reading failure—Prediction, diagnosis, intervention*. New York: Harper Row.

Johnson, M. L. (1975). The meaning of *before* and *after* for pre school children. *Journal of Experimental Child Psychology, 19*, 88-99.

Katz, R. B. (1985). Phonological deficiencies in children with reading disability: Evidence from an object-naming task. *Cognition, 22*, 225-257.

Kimball, J. P. (1973). Seven principles of surface structure parsing. *Cognition, 2*, 15-47.

Lewis, D. (1979). Scorekeeping in a language game. *Journal of Philosophical Logic, 8*, 339-359.

Liberman, A. M., Mattingly, I. G., & Turvey, M. T. (1972). Language codes and memory codes. In A. W. Melton & E. Martin (Eds.), *Coding processes and human memory*. Washington, DC: Winston and Sons.

Liberman, I. Y., & Shankweiler, D. (1985). Phonology and the problem of learning to read and write. *Remedial and Special Education, 6*, 8-17.

Liberman, I. Y., Shankweiler, D., Liberman, A. M., Fowler, C., & Fischer, F. W. (1977). Phonetic segmentation and recoding in the beginning reader. In A. S. Reber & D. L. Scarborough (Eds.), *Toward a psychology of reading: The Proceedings of the CUNY Conferences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Linebarger, M. C., Schwartz, M. F., & Saffran, E. M. (1983). Sensitivity to grammatical structure in so-called agrammatic aphasics. *Cognition, 13*, 361-392.

Lukatela, K. (1989). *Sentence processing in fluent and non-fluent aphasia*. Unpublished doctoral dissertation, University of Connecticut, Storrs.

Lundberg, I., Olofsson, A., & Wall, S. (1980). Reading and spelling skills in the first school years predicted from phonemic awareness skills in kindergarten. *Scandinavian Journal of Psychology, 21*, 159-173.

Mann, V. A., Liberman, I. Y., & Shankweiler, D. (1980). Children's memory for sentences and word strings in relation to reading ability. *Memory and Cognition, 8*, 329-335.

Mann, V. A., Shankweiler, D., & Smith, S. T. (1984). The association between comprehension of spoken sentences and early reading ability: The role of phonetic representation. *Journal of Child Language, 11*, 627-643.

Marin, O. S. M., Saffran, E. M., & Schwartz, M. F. (1976). Dissociations of language in aphasia: Implications for normal functions. *Annals of the New York Academy of Sciences, 280*, 868-884.

Marlsen-Wilson, W. D., & Tyler, L. K. (1980). The temporal structure of spoken language understanding: The perception of sentences and words in sentences. *Cognition, 8*, 1-74.

Martin, R. C. (1985). The relationship between short-term memory and sentence comprehension deficits in agrammatic and conduction aphasics. Paper presented at the Annual Meeting of the Academy of Aphasia, Pittsburgh, PA.

Martin, R. C. (1990). Neuropsychological evidence on the role of short-term memory in sentence processing. In G. Vallar & T. Shallice (Eds.), *Neuropsychological impairments of short-term memory*. Cambridge, England: Cambridge University Press.

Matthei, E.M. (1981). Children's interpretation of sentences containing reciprocals. In S. L. Tavakolian (Ed.), *Language acquisition and linguistic theory*. Cambridge, MA: MIT Press.

Mattingly, I. G., & Liberman, A. M. (in press). Specialized perceiving systems for speech and other biologically significant sounds. In G. M. Edelman, W. E. Gall, & W. M. Cowan (Eds.), *Functions of the auditory system*. New York: Wiley.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limitations of our capacity for processing information. *Psychological Review, 60*, 81-97.

Morais, J., Cluytens, M., & Alegria, J. (1984). Segmentation abilities of dyslexics and normal readers. *Perceptual and Motor Skills, 58*, 221-222.

Olson, R. K., Davidson, B. J., Kliegl, R., & Davies, S. E. (1984). Development of phonetic memory in disabled and normal readers. *Journal of Experimental Child Psychology, 37*, 187-206.

Otsu, Y. (1981). *Universal grammar and syntactic development in children: Toward a theory of syntactic development*. Unpublished doctoral dissertation, MIT, Cambridge, MA.

Perfetti, C. A. (1985). *Reading ability*. New York: Oxford University Press.

Perfetti, C. A., & Lesgold, A.M. (1977). Discourse comprehension and sources of individual differences. In M. A. Just & P. A. Carpenter (Eds.), *Cognitive processes in comprehension*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Richardson, E., & DiBenedetto, B. (1986). *Decoding skills test*. Parkton, MD: York Press.

Roeper, T. W. (1972). *Approaches to a theory of language acquisition with examples from German children*. Unpublished doctoral dissertation, Harvard University, Cambridge, MA.

Saffran, E. M. (1985). Short-term memory and sentence processing: Evidence from a case study. Paper presented at the Annual Meeting of the Academy of Aphasia, Pittsburgh, PA.

Scholes, R. J. (1978). Syntactic and lexical components of sentence comprehension. In A. Caramazza & E. . Zurif (Eds.), *Language acquisition and language breakdown: Parallels and divergences*. Baltimore, MD: The Johns Hopkins University Press.

Shankweiler, D., & Crain, S. (1986). Language mechanisms and reading disorder: A modular approach. *Cognition, 24*, 139-168.

Shankweiler, D., Crain, S., Brady, S., & Macaruso, P. (in press). Identifying the causes of reading disability. In P. E. Gough (Ed.), *Reading acquisition*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Shankweiler, D., Crain, S., Gorrell, P., & Tuller, B. (1989). Reception of language in Broca's aphasia. *Language and Cognitive Processes, 4*, 1-33.

Shankweiler, D., & Liberman, I. Y. (1972). Misreading: A search for causes. In J. F. Kavanagh & I. G. Mattingly (Eds.), *Language by ear and by eye: The relationships between speech and reading*. Cambridge, MA: MIT Press.

Shankweiler, D., Liberman, I. Y., Mark, L. S., Fowler, C. A., & Fischer, F. W. (1979). The speech code and learning to read. *Journal of Experimental Psychology: Human Learning and Memory, 5*, 531-545.

Sheldon, A. (1974). The role of parallel function in the acquisition of relative clauses in English. *Journal of Verbal Learning and Verbal Behavior, 13*, 272-281.

Smith, S. (1987). *Syntactic comprehension in reading-disabled children*. Unpublished doctoral dissertation, University of Connecticut.

Stanovich, K. E. (1982). Individual differences in the cognitive processes of reading: 1. Word decoding. *Journal of Learning Disabilities, 15*, 449-512.

Stein, C. L., Cairns, H. S., & Zurif, E. B. (1984). Sentence comprehension limitations related to syntactic deficits in reading-disabled children. *Applied Psycholinguistics, 5*, 305-322.

Tavakolian, S. L. (1981). *Language acquisition and linguistic theory*. Cambridge, MA: MIT Press.

Vallar, G., Basso, A. & Bottini, G. (1990). Phonological processing and sentence comprehension: A neuropsychological case study. In G. Vallar & T. Shallice (Eds.), *Neuropsychological impairments of short-term memory*. Cambridge, England: Cambridge University Press.

Vellutino, F. R. (1979). *Dyslexia: Theory and research*. Cambridge, MA: MIT Press.

Whitaker, H. (1976). A case of isolation of the language function. In H. Whitaker and H. A. Whitaker (Eds.), *Studies in neurolinguistics, Volume 2*. New York, NY: Academic Press.

Wingfield, A., & Butterworth, B. L. (1984). Running memory for sentences and parts of sentences: Syntactic parsing as a control function in working memory. In H. Bouma & D.G. Bouwhuis (Eds.), *Attention and performance X: Control of language processes*. Hillsdale, N.J: Lawrence Erlbaum Associates.

Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin, 101*, 192-212.

Wolf, M. (1981). The word-retrieval process and reading in children and aphasics. In K. Nelson (Ed.), *Children's language, Volume 3*. New York: Gardner Press.

FOOTNOTES

*Appears in G. Vallar & T. Shallice (Eds.), *Neuropsychological impairments of short-term memory*. Cambridge, England: Cambridge University Press (1990).

†Also University of Connecticut, Storrs.

[1]For instance, poor readers have been found to be less accurate than their age-matched controls in understanding the infinitival complements of object-control adjectives like "easy" (Byrne, 1981), which Chomsky (1969) found problematic for children as old as nine. Relative clauses also pose difficulties for poor readers in some contexts (Byrne, 1981; Goldsmith, 1980; Mann, Shankweiler, & Smith, 1984; Stein et al., 1984; but also see Smith, 1987); the relative clause is a grammatical structure which many researchers believe to develop late (Brown, 1973; deVilliers, Tager-Flusberg, Hakuta and Cohen, 1979; Sheldon, 1974; Tavakolian, 1981; but see Hamburger and Crain, 1982). Poor readers have also been shown to encounter difficulty with some dative constructions (Fletcher, et al., 1981) that have also been found to evoke comprehension errors in young children (Scholes, 1978).

[2]Further discussion of the implications of this hypothesis concerning the causes of reading disorder is presented in Crain and Shankweiler (1987), Shankweiler and Crain (1986), Shankweiler et al. (in press).

[3]This distinguishes our conception of verbal working memory from proposals by Baddeley (1986) and Carpenter and Just (1988). These researchers see working memory as a general purpose device which plays a central role not only in language, but also in reasoning, problem solving and in other forms of complex thinking.

[4]As we conceive of it, a plan is a mental representation used to guide action. The formation of a cognitive plan is an important aspect of any comprehension task involving the manipulation of objects, since the complexity of a plan is a potential source of difficulty in sentence comprehension. Therefore, when children perform poorly in language comprehension tasks that involve planning, it is important to consider the possibility that formulating the relevant plan is the locus of the problem, rather than other purely linguistic aspects of the task, such as their imperfect knowledge of linguistic rules. Where our objective is the assessment of the extent of children's linguistic knowledge (i.e., competence), we must develop our understanding of the nature and relative complexity of plans (whereby that competence is demonstrated), and we must also devise experimental paradigms in which the impact of plan complexity on linguistic processing is minimized (see Section 4 below). In addition to plan complexity, the compiling analogy allows us to entertain several other types of difficulty that a sentence can present to a subject whose task it is to plan and execute an appropriate response.

[5]Although the issue is too thorny to take up here, it may be useful to mention just one of many apparent counterexamples to the isomorphism between syntax and semantics. The example involves the phrase "three consecutive rainy days." The preferred, but noncompositional interpretation of this phrase means "three consecutive days on which it rained." The compositional interpretation requires a composite set of ordered rainy days, perhaps taken from a set of weeks, each containing a rainy day or two. So if it had rained on three consecutive Thursdays, Donald might have forgotten his umbrella on three consecutive rainy days. Since the compositional interpretation requires a more complex set of presuppositions, the default is to interpret this phrase in a way that violates compositionality. As we noted, it is reasonable to suppose that this could add to processing difficulty.

[6]There are other circumstances in which noncontiguous elements must be combined, in violation of condition (ii). This problem arises whenever there are discontinuous dependencies in the source language. In English, for example, this problem (of inverse compiling discontinuity) occurs in verb-particle constructions such as *look up*, and in cases of verb subcategorization (e.g., the verb *put* takes both an NP and a PP complement). Clearly, these properties of the linguistic input

are seen to impose demands on memory, by postponing the closure of syntactic categories.

[7]It is worth noting that changes in word order could circumvent the sequencing problem, by permitting a more nearly optimal translation between source and target language. Consider a language in which nominal modifiers appear after the noun, as in "...bear brown second." In this language, fragments of the target plan would be composed in the order in which they were mentioned in the phrase. However, it should be pointed out that in many situations, the construction of plans which satisfy condition (iii) may still require a commitment of memory resources. Specifically, the memory system must retain a data structure corresponding to the semantic value of each element of the phrase. For example, interpreting the noun "bear" as part of the phrase "bear brown second" requires the formation of a data structure delimiting the set of bears in the discourse setting. This data structure must somehow be stored, awaiting the formation of the semantic value for the remainder of the phrase. This burden on memory could be eliminated in certain contexts, such as situations in which the objects denoted by the noun are aligned in a way that allows one to restrict attention to that set alone. This would occur, for instance, if all of the bears were in the same field of view. A simple shift of gaze would allow the parser to restrict attention to the remainder of the phrase.

[8]At present, we cannot say whether unmet presuppositions thwart cognitive compiling and thereby result in premature execution or whether, instead, the satisfaction of presuppositions facilitates compile-mode behavior. In any event, it is clear that children's mistaken responses in situations that flout presuppositions exonerates the syntactic component and implicates limitations in working memory capacity.

[9]An example may be helpful in illustrating this point. It is proposed in Hamburger and Crain (1987) that certain principles of compositional semantics restrict the set of word-order possibilities in natural language. In determining semantic constituency relations, the following principle is proposed:

(P) An operator must be composed with appropriate operands.

To apply this principle, let us consider two categories of modifiers, selectors and restrictors. Modifiers like "striped" are restrictors, since they perform the semantic operation of acting on a set to produce a set as output. A selector also acts upon a set, but it returns an element as output. For example, in the phrase "second striped ball" the value of applying "second" to the set designated by "striped ball" is an individual ball. Principle (P) allows a restrictor to apply before a selector, but not after one, since the output produced by a selector is not appropriate input to a restrictor. This explains the illicit character of "striped second ball," even though this alternative phrasal order might be preferred from the standpoint of the wo.'king memory system.

[10]It is important to note that even the test sentences used in Experiment I are not the least complex case of the relevant construction, since the noun phrases in both clauses contained an inessential prenominal modifier. We chose to introduce this slight additional complexity in order to avoid inducing ceiling performance by subjects so as to make reader group differences more pronounced.

[11]In evaluating the results of this experiment, two points should be kept in mind. First, in the temporal terms experiments, there were two sources of errors: ordering errors and object selection errors. For example, in (11), an ordering error would consist of pushing the smallest horse first, whereas an object selection error might result in picking up the middle-sized horse. Unless otherwise indicated, errors of both types were combined in the analyses. Second, all of the ANOVAs we report treated subjects, and not items, as a random factor. Therefore, the findings cannot be generalized to items outside of these experiments.

[12]It should be noted that the obvious control sentences—those sentences which express agreement between order-of-mention and conceptual order—are not adequate for our purposes. Correct responses to these sentence might not be indicative of a subject's successful comprehension; rather, correct responses could reflect a simple strategy to act out complex sentences in an order-of-mention fashion. That is, they might evoke the right response for the wrong reason.

[13]It should be noted that a processing limitation account would be compatible with one kind of interaction, in which reader group differences increase to a greater extent on sentences that are particularly costly of working memory resources, such as relative clauses and 'deep' PPs. There is a hint of this type of interaction in Figure 4, and this interaction has been obtained with other populations who exhibit more severe limitations in memory capacity (e.g., Lukatela, 1989).

[14]It is worth noting that even normal adults respond incorrectly to garden path sentences in apparent violation of Subjacency when demands on working memory are raised substantially (Crain and Fodor, 1985). This supports our conclusion that the problems poor readers have with these same sentences are also due to processing factors.

[15]The two-letter code indicates how these sentence types differ. The first letter represents the noun phrase in the matrix sentence (Subject or Object) that bears the relative clause; the second letter represents the 'empty' noun phrase position in the relative clause. For example, an OS relative like (20) is one in which the Object of the main clause is modified by a relative clause with a superficially null Subject.

Haskins Laboratories Status Report on Speech Research
1991, SR-105/106, 43-52

Explaining Failures in Spoken Language Comprehension by Children with Reading Disability*

Stephen Crain† and Donald Shankweiler†

For some years our research has had the broad aim of understanding the processes whereby the language apparatus, which is biologically specialized for speech, becomes adapted to accept orthographic input. Only with such basic knowledge in hand can we hope to discover why some children fail to learn to read, and only then can we meet the challenge of effective prevention and remediation. Early work of the research group at Haskins Laboratories centered on the role of awareness of phonological segments in learning to read in an alphabetic system. It soon became evident that children who are failing to learn to read have a range of problems in the phonological sphere. In addition to problems in segmental awareness, poor readers have difficulties in naming objects, in processing speech under difficult listening conditions and are slower and less accurate in producing tongue-twisters. The cooccurrence of these problems suggested that the nature of the children's difficulty in learning to read might lie in the underlying phonological processes themselves (Liberman & Shankweiler, 1985; Liberman, Shankweiler, & Liberman, 1989).

If the fault is in phonological processes, poor readers would also be expected to have difficulties in working memory, as this form of memory relies heavily on coding based on phonological structure (Baddeley, 1966; Conrad, 1964; 1972). The memory limitations of poor readers have been noted time and again on a variety of measures that tax the phonological aspects of language processing. Some of these measure have been shown to have predictive value, singling out preschool children who will develop reading problems later. (By "poor readers" we mean those children who show a marked disparity between their measured reading skill and the level of performance that could be expected given their [normal] intelligence and opportunity for instruction. Our research compares performance by these children with age-matched controls—children who are proceeding at the expected rate in the acquisition of reading skills.)

The hypothesis that reading disability reflects a limitation in phonological processing is challenged by two findings that have emerged from the classroom as well as the laboratory. One is that some children fail to comprehend a sentence in text even when they manage to decode all the words it contains. A second challenge to a phonological explanation is the finding that some children who are poor readers fail to correctly comprehend certain sentences, in particular those with complex syntactic structure. These difficulties would seem to implicate language problems beyond the level of phonology, possibly originating in the syntactic component of language. Our recent research has taken up these challenges to the phonological explanation.

Because the question is so important, we have devoted much effort to exploring the possibility that some poor readers have underlying problems that are *not* phonological in origin. In this paper we will focus on poor readers' ability to comprehend spoken sentences. Although the findings are not wholly consistent—a few studies have failed to turn up evidence of reader group differences (e.g., Shankweiler, Smith, & Mann, 1984; Vogel, 1975)—there is evidence that good readers are significantly more accurate than age-matched poor readers in comprehension of spoken

The research discussed in this paper was supported in part by a Program Project Grant to Haskins Laboratories from the National Institute of Child Health and Human Development (HD-01994). Portions of this chapter are adapted from an earlier paper which appears in I. G. Mattingly and M. Studdert-Kennedy (Eds.), (1991), *Modularity and the motor theory of speech perception*. Hillsdale, NJ: Lawrence Erlbaum.

sentences with relative clauses, temporal terms, and adjectives with exceptional control properties (Byrne, 1981; Mann, Shankweiler, & Smith, 1984; Smith, Macaruso, Shankweiler, & Crain, 1990). The indications that poor readers do not always comprehend spoken sentences as well as good readers lend support to the possibility that limitations in phonological processing is only one of the barriers to comprehension.

It must be underscored that the comprehension difficulties noted in poor readers are typically restricted to sentences containing complex syntactic structures. From these observations it has often been inferred that the problem structures are absent or incompletely represented in the grammars of many poor readers (e.g., Byrne, 1981; Stein, Cairns, & Zurif, 1984). Since each of these structures has been claimed to be late-emerging in the course of language development, it has been proposed that many poor readers are in some sense "language delayed." Our research has sought to test a version of this hypothesis that holds that these poor readers suffer from a developmental lag in *syntactic* knowledge. We have called this the *structural lag hypothesis*.

The structural lag hypothesis has a great deal to recommend it. It explains why many children who experience difficulty learning to read also suffer in spoken language comprehension. It also explains why their comprehension difficulties are selective; poor readers should fail to comprehend those linguistic constructions that emerge late in the course of language development, but should be the equals of good readers in comprehending early-appearing constructions. As these observations suggest, the structural lag hypothesis is tied to an assumption about the course of language acquisition, and to an assumption about linguistic complexity. It supposes that certain linguistic structures develop before others, with the order of acquisition determined by the relative complexity of the structures.

The structural lag hypothesis draws support from some classical studies in language acquisition that find the late emergence of sentences with temporal terms, relative clauses, and adjectives such as *easy*, as in *The doll is easy to see* (C. Chomsky, 1969; Clark, 1970; Sheldon, 1974). In light of these findings from research on child language, the hypothesis can readily explain the observed differences between good and poor readers on spoken sentences involving these constructions. As we saw, comprehension problems are anticipated on late emerging structures that are beyond the developmental level of poor readers.

It is important to recognize, however, that by allowing at least two basic deficits in poor readers the structural lag hypothesis abandons a unitary explanation of reading disability. The limitations of poor readers in comprehension of spoken sentences are seen to be independent of their deficits in analyzing phonological information. We have proposed an alternative hypothesis that attempts to explain the entire symptom complex of poor readers, including their difficulties in spoken sentence comprehension, as a consequence of deficient phonological processing. We call this the *processing limitation hypothesis* (see also Shankweiler & Crain, 1986; Shankweiler, Crain, Brady, & Macaruso, in press).

To explain how the difficulties in understanding spoken sentences might be derived from deficient phonological processing, a few remarks are in order about our conception of the architecture of the language apparatus. Within this framework it is explained how the failures of poor readers to comprehend sentences can be directly related to their limitations in processing at the phonological level. Then we turn to the laboratory, to present evidence in support of the view that the differences between good and poor readers in spoken language comprehension are a manifestation of their differences in ability to process phonological structures.

Comprehension and the Language Apparatus

We hold the position that language processing is accomplished by a biologically coherent system in isolation from other cognitive and perceptual systems. In contemporary terms, language forms a module (Fodor, 1983; Liberman & Mattingly, 1989). We extend this notion of modularity to differentiate subcomponents of the language faculty (Forster, 1979). We see the language apparatus as composed of a hierarchy of structures and processors. The structures include the phonology, the lexicon, syntax and semantics. Each level of structure is served by a special purpose processing mechanism, or parser. A parser consists of algorithms for accessing the rules used to assign structural representations, and it may also contain mechanisms for resolving ambiguities that may arise.

We assume that the transfer of information within the language apparatus is unidirectional, beginning at the lowest-level with phonological processing and proceeding upward to the syntactic

and semantic parsers. A further assumption is that, in the course of sentence processing, the entire system works on several levels in parallel, with the operations of the various components interleaved in time, rather than in strict sequence. This permits the system to function on-line. The responsibility of synchronizing the transfer of information between levels is relegated to the verbal working memory system. Given the prominent role that this system plays in explaining the symptom complex of disabled readers, it will be worthwhile to describe our conception of working memory in slightly more detail.

The verbal working memory system. Along with other researchers, we envision the verbal working memory system as having two parts (e.g., Baddeley & Hitch, 1974; Daneman & Carpenter, 1980). First, there is a storage buffer where rehearsal of phonetically coded information takes place. This buffer has the properties commonly attributed to short-term memory: It can hold linguistic input only briefly, perhaps just for a second or two, in the order of arrival, unless the material is maintained by continuous rehearsal. The limits on capacity of the buffer mean that information must be rapidly encoded in a more durable form, beginning with phonological processing, if it is to be retained for subsequent analysis at higher levels of the language apparatus.

The second component of working memory is a control mechanism, whose primary task is to relay the results of lower-level analyses of linguistic input upward through the system. To keep information flowing smoothly, the control mechanism must avoid unnecessary computation that would stall the rapid extraction of meaning. We would speculate that the language faculty has responded to limited working memory capacity by evolving special-purpose parsing mechanisms. The parsers organize information (and resolve ambiguities), which the control component of working memory then shunts upward to the next level of the system, allowing the previous contents of the parsers to be abandoned. Rapid on-line parsing, in turn, explains how individuals with drastically-curtailed working memory capacity—capable of retaining only two or three items of unstructured material—are sometimes able to comprehend sentences of considerable length and complexity (Martin, 1985; Saffran, 1985).

To see what is most costly of memory resources, we have found it useful consider situations that are amenable to straightforward transfer of information between levels (Crain, Shankweiler, Macaruso, & Bar-Shalom, 1990; Hamburger & Crain, 1987). In the simplest case, (a) each well-formed fragment of language code at lower levels of representation is associated with a single constituent of code at higher levels, (b) the fragments of code at each level can be concatenated to form the correct representation of the input, (c) the fragments can be combined in the same order that they are accessed, and (d) each fragment is processed immediately after it is formed, permitting the source code to be discarded. These four conditions form a straightforward translation process of sequential look-up-and-concatenation familiar in the compiling of programming languages. However, all these conditions are rarely met in ordinary language. And when they are not, the computations involved in reaching the target code, for example, the semantic interpretation of a sentence, could stretch the resources of verbal working memory.

We are now prepared to show how the various difficulties manifested by poor readers can be explained in terms of the functional architecture of language. A modular view of the language apparatus raises the possibility that a deficit at the level of phonology may be the source of the entire complex of language-related deficits that characterizes reading disability. As the other features of the symptom complex can be seen as stemming from a phonologic-based deficiency, the task that remains is to explain how the difficulties that poor readers encounter in spoken language comprehension also implicate the phonological component.

Put simply, our account is as follows. We saw that the regulatory duties of working memory begin at the lowest level by bringing phonetic (or orthographic) input into contact with phonological rules, for word level analysis. In our view, this is the site of constriction for poor readers. One thing leads to another: A low-level deficit in processing phonological information creates a bottleneck that impedes the transfer of information to higher levels in the system. In other words, the constriction arises because in language processing the bottom-up flow of information from the phonologic buffer is impeded by the difficulties in accessing and processing phonological information. Therefore, all subsequent processes in the language system will be adversely affected. (Perfetti, 1985, presents a similar proposal).

Testing competing hypotheses about the source of comprehension failure

Much of our recent research has centered on testing alternative explanations of the sentence comprehension problems of poor readers. Our research strategy has two components. First, we have investigated structures that are thought to emerge late in the course of normal language acquisition. Then, for each construction we designed a pair of tasks that vary memory load while keeping syntactic structure constant. If reading disability stems from a structural lag, then children who have reading problems should perform poorly on both tasks. But according to the processing limitation hypothesis, poor readers should have greater difficulty than their age-matched controls only in tasks that place heavy demands on working memory, whatever the inherent complexity of the pertinent linguistic structures. When the same test materials are presented in tasks that minimize processing load, poor readers should do as well as good readers.

The early emergence of grammatical competence by both good and poor readers follows, in part, from our adherence to the theory of Universal Grammar. Universal Grammar maintains that many basic organizational principles of linguistic structure are innately specified (Chomsky, 1965; 1981). In keeping with the precepts of the theory, acquisition of syntactic structures seems to be essentially complete by the time instruction in reading and writing begins. The early emergence of syntax is seen to be a consequence of the innate specification of many syntactic principles that either come "prewired" or are subject to rigid system-internal constraints on grammar construction (see Crain & Fodor, 1989, for a sample of empirical research). As syntactic structures are largely built into the blueprint for language acquisition, it follows that inherent complexity of grammatical structures, as such, will not be a source of reader group differences (Crain & Shankweiler, 1988). Poor readers will be at a disadvantage, however, in contexts that stress verbal working memory.

Comprehension of temporal terms. To illustrate how we have tested the competing hypotheses, let us consider one way that linguistic input can deviate from the simple look-up-and-concatenate procedure that is hypothesized to impose minimal demands on working memory. Recall that condition (iii) of this best-case scenario would have the order of the linguistic input mirror the order in which it is composed into structural representations at higher levels. We will call a violation of this condition a *sequencing problem* (Crain, 1987). A sequencing problem arises in sentences containing temporal terms such as *before* and *after*. These terms explicitly dictate the conceptual order of events, but they may present problems of sequencing if the order in which events are mentioned conflicts with the conceptual order. This kind of conflict is illustrated in sentence (1). Note that the order in which the events are mentioned in the sentence is opposite to the order in which one would respond to the request.

(1) Push the motorcycle after you push the helicopter.

It is reasonable to suppose that sequencing problems exact a toll on the resources of working memory because both clauses must remain available long enough to enable the perceiver to formulate a response plan that represents the conceptual order. The conceptually correct response requires the formation of a two-slot template and a specification of the sequence in which the two actions are to be carried out. The information in both clauses must be held in memory long enough to put the first-mentioned action into the second slot.

There is evidence from research on language acquisition that this kind of deviation from the simple translation process is costly to working memory resources. Several studies have found that young children frequently misinterpret sentences like (1) by acting out their meanings in an order-of-mention fashion (Clark, 1970; Johnson, 1975). This response presumably reflects the simple translation process that children adopt as a default procedure for interpreting sentences that exceed their memory capacity.

An alternative explanation of the difficulties that children encounter with such sentences has been offered, however. It has been suggested that children's order-of-mention response to sentences like (1) reflect the absence in child grammar of structural knowledge that is essential to comprehension of sentences with temporal terms. This interpretation of children's errors is buttressed by the finding that they have difficulty with temporal term sentences like (1), which pose conflicts between order-of-mention and conceptual order, and not with sentences with similar meaning but with simpler syntax such as the coordinate structure sentence in (2) (Amidon & Carey, 1972).

(2) Push the motorcycle last; push the helicopter first.

However, we have questioned the assumption that sentences (1) and (2) are equivalent in meaning. Earlier studies which obtained differential responses to (1) and (2) failed to control for a presupposition that is present just in sentences like (1) (Crain, 1982; Gorrell, Crain & Fodor, 1989). The presupposition associated with this sentence is that the hearer intends to push a helicopter. To satisfy this presupposition, the subject should have established this intention *before* the command in (1) is given. A procedure which allows subjects to establish in advance their intent to perform the action mentioned in the clause introduced by the temporal term was incorporated into a study by Crain (1982). Children are asked, before each test sentence is presented, to identify one object they want to play with in the next part of the game. The experimenter subsequently incorporates this information in the subordinate clause introduced by the temporal term. For instance, sentence (1) would have been presented only after a subject had selected the helicopter, which makes the use of the temporal term felicitous.

When young children were given this contextual support, they displayed unprecedented success in comprehending sentences with temporal terms. Thus, the mistake in research that resulted in high error rates was to present sentences like (1) in the null context, which fails to satisfy the presupposition inherent in the use of temporal terms. In the null context, unmet presuppositions must be "accommodated" into the listener's mental model of the discourse setting (Lewis, 1979). Compensating for unmet presuppositions requires the hearer to revise his/her current mental model (to make it match the model of the speaker). The process of revising one's model of the discourse is seen to highly tax processing resources (see Crain & Steedman, 1985; Hamburger & Crain, 1982). If this reasoning is sound, children's grammars should be exonerated from responsibility for the errors that occurred in research that failed to satisfy the presuppositions of the test sentences.

Returning to the comprehension problems of poor readers, we saw that according to the processing limitation hypothesis their performance should suffer appreciably in contexts that tax working memory. It seems reasonable to suppose, therefore, that poor reader's special limitations in working memory would cause them to have greater difficulty than good readers in processing sentences containing temporal terms *if they are presented in the null context*. However, the processing limitation hypothesis would anticipate that both good and poor readers would display a high rate of successful comprehension in felicitous contexts. The structural lag hypothesis, on the other hand, would anticipate the same differences between good and poor readers both with and without contextual support, because lightening the burdens imposed on working memory should not result in improved comprehension of a structure that is absent from a child's internal grammar.

We investigated these contrasting predictions in a figure-manipulation task in which sentences with temporal terms were auditorily presented to good and poor readers (Macaruso, Bar-Shalom, Crain, & Shankweiler, 1989). Our experiment was designed to exacerbate the processing load on both reader groups by including an additional prenominal modifier in half of the test sentences. As exemplified in (3), the main clause of these sentences contained complex NPs with an ordinal quantifier (*second*) and a superlative adjective. Adjectives combine to introduce added complexity to the plan that one must formulate in order to respond accurately to the sentence. The remaining test sentences contained "simple NPs", i.e., with no additional ordinal modifier in the main clause, as in (4).

(3) Push the *second smallest horse* before you push the blue car.
(4) Pick up the *largest truck* after you pick up the blue horse.

The stimuli consisted of 16 sentences with temporal terms *before* and *after*. Four sentences were presented in which the order-of-mention of events was the same as the conceptual order, as in (3). In the remaining twelve sentences, the order of mention was opposite to the conceptual order, as in (4). Children encountered the test sentences in two contexts: one that satisfied the presupposition associated with the use of temporal terms, and one that did not (the null context).

As anticipated, poor readers performed less well overall than good readers in acting out sentences containing temporal terms. However, by satisfying the felicity conditions and thereby reducing memory demands, we obtained a significant reduction in errors for both groups combined. Moreover, the satisfaction of presuppositions benefited poor readers more than good readers. This lends credence to the hypothesis that, without contextual support, poor readers' limitations in working memory are exacerbated. However, the poor readers performed at a success rate of 82.4% when the felicity conditions were satisfied, even when half

of the test sentences contained complex NPs. This calls into question the claim of the structural lag hypothesis that poor readers lag in their mastery of complex syntactic structures.

Further support for a processing interpretation of poor readers' comprehension difficulties comes from the finding that poor readers were more adversely affected by changes in NP complexity than good readers. The special problems that poor readers had with the complex NP sentences presumably reflect the fact that these sentences are more taxing on working memory resources, as discussed earlier.

Additional evidence of processing difficulty was obtained when we compared responses to the two types of sentences that present a conflict between order-of-mention and conceptual order. Notice that the presence of a temporal term in the initial clause eases the burden on working memory, by indicating in advance that a two-slot template is required, as in (5). Here, the use of *before* in the initial clause delays execution. This contrasts with the corresponding sentences with *after*, such as (6), where the temporal conjunction is contained in the second clause.

(5) Before you push the helicopter, push the motorcycle.
(6) Push the motorcycle after you push the helicopter.

On the account of working memory that we have proposed, we would predict the sentences with *after* to be harder, since the subject has no warning that information should be maintained in memory while awaiting subsequent material. As expected, poor readers were least successful in responding to *after* sentences that presented a conflict between order-of-mention and conceptual order and no contextual support in the form of satisfied presuppositions. Good readers, on the other hand, were not as handicapped by the memory demands imposed by these sentences.

Taken together, the findings of the experiment by Macaruso et al. indicate that, as processing demands are increased, poor readers' performance involving temporal terms sentences is eroded much more than good readers' performance. Decreasing processing demands, either by satisfying the felicity conditions or by using less complex NPs, elevates performance by poor readers, such that group differences diminish. In the best case, both reader groups perform at a high level of success.

Comprehension of garden path sentences. Another way to address the question of a processing limitation versus a structural deficit is to examine the pattern of errors across constructions for each reader group. A processing limitation, and not a structural deficit, can be inferred if both reader groups reveal a similar pattern of errors across sentence types. Pursuing this research strategy, another study (Crain et al., 1990) asked how good and poor readers would respond to the kind of garden path sentences that are created when listeners follow a parsing strategy for resolving structural ambiguities, called *Right Association* by Kimball (1973) and *Late Closure* by Frazier (1978).

Late Closure encourages listeners or readers to connect an incoming phrase as low as possible in the phrase marker that has been assigned to the preceding material. It seems reasonable to suppose that this parsing strategy reflects the functional architecture of the language apparatus, which has many computations to perform and little available space for their compilation and execution. Although strategies such as this may have evolved to enable the parser to circumvent the limitations of working memory, they may introduce new problems of their own, because the decision dictated by a strategy may turn out to be incorrect in light of subsequent input. Clearly, recovery from these so-called garden paths is possible only within the limits of working memory, because these limits determine whether the grammatically correct analysis is still available. Because sentences that tax working memory heavily have been found to present special difficulties for poor readers, they should be less able than good readers to recover from garden paths prompted by Late Closure.

We tested this prediction by asking good and poor readers to respond to several types of garden path sentences. In each of these, the parse favored by Late Closure tempts one to make an ungrammatical analysis, in which the extraction of the *Wh*-phrase violates a putatively innate constraint called *Subjacency*. Subjacency establishes the boundary conditions on the movement of Wh-phrases in the formation of questions. Specifically, it prohibits movement over more than a single "bounding node" (NP or S in English). One consequence of Subjacency is that Wh-phrases cannot be extracted out of complex NPs like those in the test sentences in this study, which were taken from Crain and Fodor (1985).

Three types of garden-path sentences were created. Each type varied in the severity of the processing load. The subsequent examination of the error pattern by the two reader groups across

sentence types was used to distinguish between the competing hypotheses about the source of comprehension difficulties in poor readers. The different syntactic constructions are illustrated in (7)-(9).

(7) Prepositional Phrase (deep): What is Jennifer drawing a picture of a boy with?
(8) Prepositional Phrase (distant): Who is Susan handing over the big heart-shaped card to?
(9) Relative Clause: Who is Bill pushing the cat that is singing to?

Sentence (7) is labeled "deep" because the origin of the "extracted" Wh-phrase is a prepositional phrase that is embedded in an NP which is itself embedded in an NP. This contrasts with the "distant" case (8), in which there is only one level of embedding. Although the sentences are matched for length, we anticipated that distant prepositional phrase sentences would be easier to process than either the deep prepositional phrase sentences or relative clause sentences like (9). The depth of syntactic embedding in both relative clause and deep prepositional phrase sentences means that they deviate more than the distant sentences from the simple look-up-and-concatenate translation process.

On each trial subjects were asked to listen carefully to a tape-recorded set of sentences that described a scene depicted in a large cartoon drawing placed in front of them. Immediately following the description, they were asked to respond to a question about some aspect of the drawing. For example, the context sentences in (10) preceded the test question (9).

(10) Bill's father is waiting for Bill to bring him the cat. The cat loves to sing and has made up a song for his toy mouse.

The grammatically correct response to this question is "his father." The response "the mouse" is incorrect, because it represents an apparent violation of Subjacency. The processing limitation hypothesis would argue that an examination of the pattern of errors across sentence types for each group may provide evidence that these errors are not actually violations of Subjacency; instead they are the result of the processing burdens these sentence impose on working memory. The structural lag hypothesis makes no definite predictions about the pattern of responses by good and poor readers for any of the sentence types presented in this experiment. It seems reasonable to suppose, however, that under this hypothesis we might anticipate a different response profile for good and poor readers, as these sentences incorporate exceedingly complex structures.

To reiterate, the processing limitation hypothesis predicts that both groups will manifest a similar pattern of errors across sentences of varying syntactic types, with poor readers penalized to a greater degree than good readers on sentences that are most costly to working memory resources (e.g., the deep prepositional phrase and relative clause sentences). This is exactly what we found. There was a general decrement in performance by poor readers, but both good and poor readers responded in a similar way to different linguistic constructions (see Figure 1).

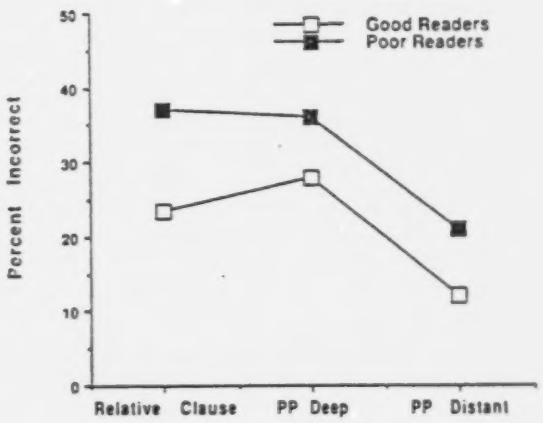


Figure 1.

Detection and correction of ungrammatical sentences. An experiment by Fowler (1988) deserves mention as another significant effort to disentangle structural knowledge and processing capabilities in beginning readers. Preliminary tests were administered to assess children's phonological awareness, working memory and spoken sentence understanding. As in earlier work, there were clear-cut correlations with measures of reading. But in addition, children were compared on a grammaticality judgement task and a sentence correction task. The judgement task is presumed to place minimal demands on working memory, so it was used to establish a baseline of the subjects' structural knowledge, for subsequent comparison with the correction task. The expectation that grammaticality judgments do not stress working memory is motivated in part by recent research

showing that, despite severe memory limitations, agrammatic aphasic patients are able judge the grammaticality of sentences of considerable length and syntactic complexity (Linebarger, Schwartz & Saffran, 1983; Saffran, 1985; Shankweiler, Crain, Gorrell & Tuller, 1989). These findings suggest that this task directly taps the syntactic analysis being assigned. In the correction task, subjects were asked to change ungrammatical sentences (taken from the judgment task) to make them grammatical. Clearly, correcting grammatical anomalies requires the ability to hold sentences in memory long enough for reanalysis.

According to the processing limitation hypothesis, both good and poor readers should do equally well on the grammaticality judgment task, but differences should occur on the correction task. This is exactly what Fowler found. Reading ability was significantly correlated with success on the correction task, but not with success on the judgment task. This is further support for the view that processing complexity, and not structural complexity, is a better diagnostic of reading disability. Two additional findings bear on the competing hypotheses about the causes of reading failure. First, the level of achievement on grammaticality judgements was well above chance for both good and poor readers, even on complex syntactic structures (e.g., Wh-movement, and tag questions). Second, results on a test of short-term recall (with IQ partialed out) were more strongly correlated with success on the sentence correction task than with success on the judgment task.

Conclusion

The manner in which reading is erected on preexisting linguistic structures led us to predict that the causes of reading disability would lie within the language domain. Accordingly, seemingly normal school children who fail to make the expected progress in learning to read were found to have language-related difficulties, including problems in phonological awareness and unusual limitations in verbal working memory. As both of these problems are arguably grounded in phonology, a central concern has been to determine if all the language-related difficulties evinced by poor readers might stem from a single deficit in processing phonological information.

The observation that poor readers have difficulties in correctly interpreting some spoken sentences seemed, at first, to threaten a unitary phonological deficit account. However, in the context of our assumptions about the architecture of the language apparatus we argued that a phonological deficit might explain this problem too. If so, this argues against attributing the comprehension difficulties of reading disabled children to a developmental lag in structural competence over and above their well-attested deficiencies in phonological processing.

In order to tease apart the alternatives, two research strategies were implemented. In one, tasks were devised that stress the language processing system in varying degrees, while holding syntactic structure constant. We reviewed an experiment that followed this strategy, which yielded large differences between good and poor readers in comprehending spoken sentences in contexts that stress working memory, but much smaller differences when the same materials were presented in a way that lessened memory load. Contrary to the expectations of the structural lag hypothesis, in contexts that minimized memory demands, both reader groups achieved such a high level of accuracy that competence with the construction under investigation would seem guaranteed.

A second research strategy tested for differences between reader groups by comparing performance across a variety of linguistic structures. As anticipated by the processing limitation hypothesis, we found that both reader groups manifest a similar pattern of errors across sentences of varying syntactic types, with poor readers penalized to a greater degree than good readers on sentences that are costly to working memory resources. The absence of a reader group interaction invites the inference that the relatively inferior performance of poor readers is due to parsing pressure, rather than to ignorance of Subjacency, a putatively innate constraint on syntax.

In sum, the syntactic component of the language apparatus appears to be intact in poor readers. The source of difficulties that might appear to reflect a syntactic deficiency must be sought elsewhere. It is premature at present to exclude the possibility that the comprehension problems of some poor readers are caused by a deficiency in some other component of the language apparatus (e.g., syntactic parsing). However, we can appeal to the modular architecture of the language apparatus to explain how a deficit in phonologic processing may masquerade as a complex of deficits throughout the whole language system. Given the abundance of evidence attesting to poor readers' deficits in the phonological domain, there is reason to prefer the hypothesis that their

comprehension problems are part and parcel of their difficulties in phonological processing. If this is correct, it would prove *unnecessary* to postulate additional impairments within the language system: all of the problems associated with reading ultimately spring from the same source. The possibility of providing a unitary explanation of an apparently disparate set of phenomena is a compelling reason, in our view, for adhering to a modular conception of the language apparatus.

REFERENCES

Amidon, A., & Carey, P. (1972). Why five-year-olds cannot understand before and after. *Journal of Verbal Learning and Verbal Behavior, 11*, 417-423.

Baddeley, A. D. (1966). Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *Quarterly Journal of Experimental Psychology, 18*, 362-365.

Baddeley, A.D. & Hitch, G. B. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8). New York: Academic Press.

Byrne, B. (1981). Deficient syntactic control in poor readers: Is a weak phonetic memory code responsible? *Applied Psycholinguistics, 2*, 201-212.

Chomsky, C. (1969). *The acquisition of syntax in children from 5 to 10*. Cambridge, MA: MIT Press.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris Publication.

Clark, E. V. (1970). How young children describe events in time. In G. B. Flores d'Arçais and W. J. M. Levelt (Eds.), *Advances in psycholinguistics* (pp. 275-284). Amsterdam: North-Holland.

Conrad, R. (1964). Acoustic confusions in immediate memory. *British Journal of Psychology, 3*, 75-84.

Conrad, R. (1972). Speech and reading. In J. F. Kavanagh and I. G. Mattingly (Eds.), *Language by ear and by eye: The relationships between speech and reading* (pp. 205-240). Cambridge, MA: MIT Press.

Crain, S. (1982). Temporal terms: Mastery by age five. In Papers and Reports on Child Language Development, Proceedings of the Fourteenth Annual Stanford Child Language Research Forum (pp. 33-38). Stanford, CA: Department of Linguistics, Stanford University.

Crain, S. (1987). On performability: Structure and process in language understanding. *Clinical Linguistics and Phonetics, 1*, 127-145.

Crain, S., & Fodor, J. D. (1985). On the innateness of Subjacency. In The Proceedings of the Eastern States Conference on Linguistics (Vol. I., pp. 191-204). Columbus, OH: The Ohio State University.

Crain, S., & Fodor, J. D. (1989). Competence and performance in child language. In E. Dromi (Ed), *Language and cognition: A developmental perspective*. Norwood, NJ: Ablex.

Crain, S., & Shankweiler, D. (1988). Syntactic complexity and reading acquisition. In A. Davison & G. M. Green (Eds.), *Linguistic complexity and text comprehension: Readability issues reconsidered* (pp.167-192). Hillsdale, NJ: Erlbaum.

Crain, S., Shankweiler, D., Macaruso, P., & Bar-Shalom, E. (1990). Working memory and sentence comprehension: Investigations of children with reading disorder. In G. Vallar & T. Shallice (Eds.), *Neuropsychological impairments of short-term memory* (pp. 477-508). Cambridge, England: Cambridge University Press.

Crain, S., & Steedman, M. (1985). On not being led up the garden path: The use of context by the psychological syntax processor. In D. Dowty, L. Karttunen & A. Zwicky (Eds.), *Natural language parsing* (pp. 320-358). Cambridge, England: Cambridge University Press.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*, 50-466.

Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.

Forster, K. (1979). Levels of processing and the structure of the language processor. In W. E. Cooper & E. Walker (Eds.), *Sentence processing: psycholinguistic studies presented to Merrill Garrett* (pp. 27-85). Hillsdale, NJ: Lawrence Erlbaum Associates.

Fowler, A. E. (1988). Grammaticality judgments and reading skill in grade 2. *Annals of Dyslexia, 38*, 73-94.

Frazier, L. (1978). *On comprehending sentences: Syntactic parsing strategies*. Unpublished doctoral dissertation, University of Connecticut.

Gorrell, P., Crain, S., & Fodor, J. (1986). Contextual information and temporal terms. *Journal of Child Language, 16*, 623-632.

Hamburger, H., & Crain, S. (1982). Relative acquisition. In S. Kuczaj (Ed.), *Language development Vol. 1: Syntax and Semantics* (pp. 245-274). Hillsdale, NJ: Lawrence Erlbaum Associates.

Hamburger, H., & Crain, S. (1987). Plans and semantics in human processing of language. *Cognitive Science, 11*, 101-136.

Johnson, M.L. (1975). The meaning of *before*. and *after* for preschool children. *Journal of Experimental Child Psychology, 19*, 88-99.

Kimball, J. P. (1973). Seven principles of surface structure parsing. *Cognition, 2*, 15-47.

Lewis, D, (1979). Scorekeeping in a language game. *Journal of Philosophical Logic, 8*, 339-359.

Liberman, A. M., & Mattingly, I. G. (1989). A specialization for speech perception. *Science* 243, 489-494.

Liberman, I. Y., & Shankweiler, D. (1985). Phonology and the problems of learning to read and write. *Remedial and Special Education 6*, 8-17.

Liberman, I. Y., Shankweiler, D., & Liberman, A. M. (1989). The alphabetic principle and learning to read. In D. Shankweiler & I. Y. Liberman (Eds.), *Phonology and reading disability: Solving the reading puzzle*. IARLD Research Monograph Series. Ann Arbor: University of Michigan Press.

Linebarger, M., Schwartz. M., & Saffran, E. M. (1983). Sensitivity to grammatical structure in so-called agrammatic aphasia. *Cognition 13*, 361-392.

Macaruso, P., Bar-Shalom, E. Crain, S., & Shankweiler, D. (1989). Comprehension of temporal terms by good and poor readers. *Language and Speech, 32*, 45-67.

Mann, V. A. Shankweiler, D., & Smith, S. T. (1984). The association between comprehension of spoken sentences and early reading ability: The role of phonetic representation. *Journal of Child Language, 11*, 627-643.

Martin, R. C. (1985). The relationship between short-term memory and sentence comprehension deficits in agrammatic and conduction aphasics. Paper presented at the annual meeting of the Academy of Aphasia, Pittsburgh, PA.

Perfetti, C. A. (1985). *Reading ability*. New York: Oxford University Press.

Saffran, E. M. (1985). Short-term memory and sentence processing: Evidence from a case study. Paper presented at Academy of Aphasia, Pittsburgh, PA.

Shankweiler, D., & Crain, S. (1986). Language mechanisms and reading disorders: A modular approach. *Cognition, 24*, 139-168.

Shankweiler, D., Crain, S., Gorrell, P., & Tuller, B. (1989). Reception of language in Broca's aphasia. *Language and Cognitive Processes*.

Shankweiler, D., Crain, S., Brady, S., & Macaruso, P. (in press). Identifying the causes of reading disability. In P. B. Gough (Ed.), *Reading acquisition*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Shankweiler, D., Smith, S. T., & Mann, V. A. (1984). Repetition and comprehension of spoken sentences by reading-disabled children. *Brain and Language 23*, 241-257.

Sheldon, A. (1974). The role of parallel function in the acquisition of relative clauses in English. *Journal of Verbal Learning and Verbal Behavior, 13*, 272-281.

Smith, S. T., Macaruso, P., Shankweiler, D., & Crain, S. (1990). Syntactic comprehension in young poor readers. *Applied Psycholinguistics, 11*.

Stein, C. L., Cairns, H. S., & Zurif, E. B. (1984). Sentence comprehension limitations related to syntactic deficits in reading-disabled children. *Applied Psycholinguistics, 5*, 305-322.

Vogel, S. A. (1975). Syntactic abilities in normal and dyslexic children. Baltimore, MD: University Park Press.

FOOTNOTES

*In D. Balota, G. B. Flores d'Arçais and K. Rayner (Eds.), *Comprehension processes in reading*. Hillsdale, NJ: Lawrence Erlbaum Associates (1990).

†Also University of Connecticut, Storrs.

Haskins Laboratories Status Report on Speech Research
1991, SR-105/106, 53-64

How Early Phonological Development Might Set the Stage for Phoneme Awareness*

Anne E. Fowler†

A recent chapter by Treiman and Zukowski (in press) provides evidence that pre-readers are aware of sublexical units intermediate between the syllable and the phoneme. Whereas four-year old preschoolers can note commonalities in sound only when a full syllable is shared (e.g., en*treat*/re*treat*), five-year old kindergartners also demonstrate an ability to group words on the basis of a shared onset or a shared rime (e.g., *treat*/*trick*, *sack*/*black*). In their study, the ability to group words on the basis of a single shared segment (e.g., *b*reak/*b*lock) does not appear before the first grade. These findings support Treiman's view, expressed in other papers (Treiman, 1985, Treiman & Danis, 1988), that syllables, onsets, and rimes constitute units of linguistic processing that are more accessible than the phoneme. In demonstrating the developmental priority of onsets and rimes, Treiman and Zukowski suggest that this initial cut may serve as a guide in focussing explicit attention on the internal structure of the syllable.

In my discussion, I would like to explore the possibility that the developmental progression observed by Treiman and Zukowski over the preschool years may extend beyond phonological awareness to reflect more fundamental changes in phonological representation, in particular to how lexical items are stored for recognition and retrieval. That is, the child's early vocabulary may originally be represented at a more holistic level, with organization in terms of phonemic segments emerging only gradually in early childhood. This suggestion is consistent with the view adopted by Treiman in earlier papers (Treiman & Baron, 1981; Treiman & Breaux, 1982) and has been supported by a number of researchers studying early phonological development (Ferguson, 1986; Jusczyk, 1986; Menyuk & Menn, 1979; Studdert-Kennedy, 1986, 1987). Because segmental organization would seem to be an essential substrate for the emergence of phonemic awareness, it may be that extending our investigation to basic phonological development will provide a more complete account of the ontogenesis of phonological awareness and of reading disability than has been available to date.

The traditional view of phonological development has suggested that the lexical structures of the child are organized in terms of phonemic segments well before awareness of these structures is required for alphabetic literacy. Accordingly, the difficulty in achieving awareness of phonemes was attributed to the metacognitive requirements of the task. In the first part of my discussion, I examine the consequences and limitations of a strictly metacognitive hypothesis in accounting for the emergence of phoneme awareness and in explaining individual differences in achieving success. In the second section, I introduce some recent evidence from research on early phonological development that may challenge this assumption, suggesting instead that lexical representations become increasingly segmental between one and eight years of age. In the third section, I examine the implications of this new perspective for explaining phenomena of phonological awareness and reading disability that are not readily handled by the metacognitive account. Finally, I take up the issue of causality, comparing two explanations for developmental

I am grateful to my many colleagues at Haskins (Catherine Best, Catherine Browman, Lois Dreyer, Alice Faber, Carol Fowler, Louis Goldstein, Eliza Goodell, Alvin Liberman, Isabelle Liberman, Susan Nittrouer, and Donald Shankweiler) for sharing their expertise and providing interesting discussion. I am particularly indebted to Michael Studdert-Kennedy and Susan Brady for their careful and insightful reading of an earlier draft of this paper. This work was done during the author's tenure as a National Down Syndrome Society scholar.

and individual differences in phonological awareness and phonological representation.

I. THE METACOGNITIVE ACCOUNT OF PHONEME AWARENESS AND READING DISABILITY

Thanks in part to Isabelle Liberman, it is now an established fact that successful reading in an alphabetic system entails phoneme awareness. Whether one wishes to argue that phoneme awareness is a precursor (Bradley & Bryant, 1983; Lundberg, Frost, & Petersen, 1988), a corequisite (Liberman, Shankweiler, & Liberman, 1989; Perfetti, Beck, Bell, & Hughes, 1987), or a byproduct of reading acquisition (Ehri, 1989; Morais, Cary, Alegria, & Bertelson, 1979), novel words cannot be decoded without concomitant awareness of the segmental nature of speech (see Gough & Walsh, in press). It is also a fact that phoneme awareness is not a necessary outcome of learning to speak a language: phoneme awareness will not ordinarily develop without specific tuition and even then rarely before five or six years of age (A. Liberman, 1989). What is not known are the developmental precursors to phoneme awareness and the reasons why success is so variable. Why is it that some children fail to perform adequately on measures of phoneme awareness, and fail to read, despite years of instruction, whereas others require but a cursory introduction for success?

In accounting for the absence of phoneme awareness in young children, investigators originally focussed on the explicit level of knowledge required (Gleitman & Rozin, 1977; I. Liberman, 1973). That is, it was assumed that preschoolers lacking phoneme awareness should nonetheless display evidence for phoneme-level organization on other phonological tasks such as perception, production, word recognition, and memory. This assumption was based on a traditional model of phonemes and features (e.g., Chomsky & Halle, 1968), and was fueled by the findings that pre-linguistic infants can discriminate virtually all phonemic contrasts (e.g., Eimas, Siqueland, Jusczyk, & Vigorito, 1971). Early research on phonological development did not contradict this assumption. On the view prevalent at that time, it was assumed that by the time word production begins, perceptual categorizations, and hence underlying representations, were equivalent to adult surface phonemic forms. Where deviations occurred in production, phonological "rules" such as reduction, deletion or substitution, were invoked to relate the production to the assumed underlying form (e.g., Smith, 1973).

Yet young children typically lack conscious awareness of phonemic segments. It is now well documented that preschoolers cannot tell you that *pat* has three separate sounds (Liberman, Shankweiler, Fischer, & Carter, 1974), produce "just a little bit" of *man* (Fox & Routh, 1976), or "say *pat* without the /p/" (Rosner & Simon, 1971). Four- and five-year olds also appear to be unable to group words together on the basis of a single common phoneme; they fail to observe the commonality between *prince* and *plate* or *sad* and *bud* (e.g., McLean, Bryant, & Bradley, 1987; Treiman & Zukowski, in press). To account for the discrepancy in young children between the knowledge of phonemes presumed to underlie their early perception and production abilities, and the lack of knowledge exhibited in segmentation and categorization tasks, a distinction was made between implicit and explicit knowledge. It was argued that gaining access to these segments in order to count, label or manipulate them for the purpose of reading is akin to becoming aware of the many movements that go into walking for the purpose of learning ballet. These are metacognitive tasks imposed upon an autonomously functioning system. (See Rozin, 1975, for an interesting discussion of the distinction between identifying the units which guide an organism's behavior and granting that organism conscious access to those units).

Explicit awareness requires that we suspend the normal function of a behavior (here, listening to speech to gain access to meaning) to focus on its formal attributes. This ability to shift readily from one aspect of a stimulus to another, often termed "decentering," is considered a hallmark of the Piagetian stage of concrete operations which begins between five and seven years of age, just when phoneme awareness is emerging. The hypothesis that metacognitive factors may account for differences in achieving phoneme awareness was given support through a number of studies demonstrating that skill in tasks requiring awareness of phonemes is associated with skill on other metalinguistic tasks involving awareness of morphological or syntactic structures (e.g., Fowler, 1988). Although the evidence is weaker when metacognitive ability outside of language is assessed, it has been suggested that children must attain some minimal threshold of cognitive development before they can grasp and refer to abstract concepts of word, syllable or phoneme. For example, children who performed poorly on a

Piagetian measure of concrete operations in kindergarten had more difficulty learning to read in first grade than children who did better (Tunmer, 1988).

Although general metacognitive factors do appear to play a role in allowing phoneme awareness to first develop, they are less successful in explaining why some individuals continue to experience difficulty, even after they have attained concrete operations and explicit attention has been called to the phonemic level through reading instruction (e.g., Bradley & Bryant, 1978; Pratt & Brady, 1988; Read & Ruyter, 1985). In fact, reading-disabled individuals from kindergarten to adulthood can perform non-linguistic metacognitive tasks involving angles or figures, but cannot perform the same task operations when the items to be accessed or manipulated are phoneme segments (Fowler, 1990; Lundberg, Olofsson, & Wall, 1980; Mann, Tobin, & Wilson, 1987).

Even awareness of larger linguistic units is not sufficient for phoneme awareness. As demonstrated quite clearly in the study by Treiman and Zukowski, long before children reveal awareness at a segmental level, it is evident that they are attending to the sound structure of language and making associations between words on the basis of a common syllable or portion of a syllable (see also Walley, Smith, & Jusczyk, 1986). Similarly, although extensive training efforts prior to age five have been successful in teaching children to segment or categorize on the basis of words or syllables, these training programs have not instilled awareness at a phoneme level, despite intensive efforts (Content, Kolinsky, Morais, & Bertelson, 1986; Fox & Routh, 1976; Treiman & Breaux, 1982). What distinguishes the younger child from the older child, or the non-reader from the successful reader, is the specific failure to access the phoneme. (See also McLean et al., 1987).

Finally, it should be noted that there is not a strong association between reading and general intelligence, especially in young children who have not yet experienced the effects of reading failure. As argued by Stanovich (1988), a failure to perform in an analytic manner should cut across all domains, yet the very hallmark of specific reading disability is a failure to learn to read despite adequate intelligence. Specific deficits in phoneme awareness continue to be associated with reading difficulty even when differences in general intelligence have been controlled for. (See Stanovich, Cunningham, & Feeman, 1984, for review and discussion).

In contrast to a strictly metacognitive perspective, there is growing evidence that the phonological problems of poor readers often extend beyond the level of awareness (e.g., Brady, 1986, in press; Liberman & Shankweiler, 1985; Stanovich, 1985; Wagner & Torgesen, 1987). Memory, perception, articulation, and lexical access have been implicated in reading disability; all ultimately depend upon phonological representations, yet none obviously requires phoneme awareness. The best documented of these areas is verbal short-term memory: poor readers are less able than better readers to retain strings of words, digits or other material that can be verbally encoded. That the difficulty is fundamentally phonological is indicated both by analysis of errors produced and by the lack of difference between good and poor readers on nonverbal memory tasks (e.g., Brady, Mann, & Schmidt, 1987; Katz, Shankweiler, & Liberman, 1981). (See Brady, in press; Jorm, 1983, and Torgesen, 1978 for evidence and discussion).

There is some evidence that reading ability is also associated with naming, the ability to rapidly and accurately produce the phonological labels of items known to be in an individual's recognition vocabulary. For example, kindergarten children who perform poorly on a task involving rapid successive naming of common pictured objects presented in an array are at risk for later reading failure (Blachman, 1984; Denckla & Rudel, 1976; Wolf, 1986). Furthermore, poor readers at a variety of ages make more errors than good readers in their attempts to produce names of objects they can recognize, producing forms phonologically related to the target item. This has been interpreted as indicating that their stored lexical representations may be less precisely specified (Catts, 1986; Katz, 1986; Wolf & Goodglass, 1986).

The third area of phonological difficulty associated with reading disability concerns speech perception and production. Contrary to the expectations of a metacognitive hypothesis, several studies have found that poor readers have difficulty with the categorical perception of certain speech contrasts (Godfrey, Syrdal-Laskey, Millay, & Knox, 1981; Pallay, 1986; Werker & Tees, 1987). Similarly, reading disabled children make significantly more perceptual errors than good readers when asked to repeat words presented in noisy listening conditions, although the two groups performed equivalently on a non-verbal control task (Brady, Shankweiler, & Mann, 1983). Poor readers also appear to be less able to accurately produce tongue twisters or to repeat

phonologically complex or unfamiliar lexical items, suggesting that their production skills may also be compromised (Catts, 1986; Rapala & Brady, 1990; Snowling, 1981; Snowling, Goulandris, Bowlby, & Howell, 1986).

What we have learned from the literature on reading disability, then, is that a failure to gain access to phonemic segments is associated not with general metacognitive inadequacies, but with a host of other subtle phonological deficits, involving the formation, retrieval and maintenance of phonological representations. Although the difficulties have been attributed variously to "weak," "fragile," or "underspecified" representations, arrived at via "inefficient" phonological processing, there has been little attempt to further define these terms or to reconcile these individual differences within the current theory of phonological development.

II. ON THE EMERGENCE OF THE PHONEME

As we noted earlier, the original concept of phoneme awareness rests crucially on the assumption that the very young child perceives, produces, and represents speech in terms of phoneme categories (Gleitman & Rozin, 1977; Liberman, 1973). More recently, however, several lines of evidence have converged to place this assumption in some doubt, suggesting that basic phonological representations undergo growth and change in the early stages of language development (Ferguson, 1986; Jusczyk, 1986; Menyuk & Menn, 1979; Studdert-Kennedy, 1986, 1987). Most critical is the suggestion that a child's first words are not, as is often assumed, represented as a sequence of independent phonemes, but may instead be stored and retrieved as a holistic pattern of interacting elements, variously described as gestures, features, or articulatory routines (Ferguson & Farwell, 1975; Menyuk & Menn, 1979). According to this view, it is only with increasing pressure of the expanding lexicon that the scope of the representation needs to be narrowed, giving salience first to the syllable, and then to subsyllabic units. The implication is that the addition of a phonemic level of organization emerges only gradually in early childhood.

This view of phonological development rests on the assumption that it is not phonemes, but features, or articulatory gestures, that are the fundamental units of perception and production (Browman & Goldstein, 1986, in press; Tartter, 1986).[1] Positing gestures as the basic components of the syllable is not only consistent with the research on infant perceptual abilities (e.g., Eimas et al., 1971; Kuhl, 1987), but also makes it possible to move smoothly from infant abilities (both perception and babbling) to first words without having to invoke phoneme representations at some intermediate point. (Studdert-Kennedy, 1986). What may be changing over the course of phonological development is the ability of the child to coordinate gestures which initially extend the full length of the syllable into integrated subsyllabic routines that recur in many different environments (Nittrouer, Studdert-Kennedy, & McGowan, 1989). As outlined by Jusczyk (1986), the scope of the gesture may narrow first to the onset/rime distinction and only later to the level of phonemic segments, thus paralleling the developmental progression observed for awareness by Treiman and Zukowski in the paper under discussion.

This view is also consistent with recent developments in phonological theory, where it has become increasingly evident that a number of linguistic phenomena characterizing a mature system make reference to the entire syllable and cannot be neatly tied to one or another phonemic segment. Such is the case, for example, for tone in Chinese or for intonation and stress patterns in English (cf. Clements & Keyser, 1983). The hierarchical models characterizing current phonology appear to instantiate the very levels postulated as occurring developmentally (e.g., Fudge, 1987). Examination of adult speech errors also provides support for non-phoneme syllable divisions in language processing, indicated by the finding that not all phoneme segments exchange at a similar rate. For example, syllable initial consonant clusters as in *break* or *stop* frequently move together as a unit (e.g., Treiman & Danis, 1988), as do vocalic nuclei composed of a vowel and a following liquid as in *hard* or *cold* (e.g., Shattuck-Hufnagel, 1987; Treiman, 1984).

Of course, from the perspective of acquiring phoneme awareness and reading, the important question pertains to when in early childhood the phoneme level of organization is sufficiently well developed to allow for the isolation, labeling, and manipulation of these segments. Although the bulk of the evidence for the phoneme as an emerging entity derives from studies focussing on early word production in the second year of life, the little evidence we have available suggests that the scope of gestures continues to become increasingly phonemic between three and seven years of age, inviting comparison with phoneme awareness abilities over the same period. In the

rest of this section, I review the kind of data available from early production and then go on to discuss the evidence for growth extending to the onset of reading acquisition.

When the assumption of the phoneme as a guiding category is suspended, it appears that during the first 50 words or so the basic unit of production is the whole word shape (Ferguson, 1986). That is, children's earliest productions are represented as distinct holistic shapes, with prosodic and articulatory attributes that are not systematically related to other words. The evidence against phoneme-level representation at this stage is generally considered to include three main points. First, phonetic forms appear to be tied to particular utterances: a phonetic form observed in one word is often not generalized across words in early productions. For example, a 15-month old girl studied by Ferguson and Farwell (1975) produced [n] correctly in *no*, but substituted [m] for [n] in *night* and [b] for [m] in *moo*. As noted by Nittrouer et al. (1989), "the child does not contrast [b], [m], and [n] as in the adult language, but the three words with their insecurely grasped onsets" (p.120). Second, production of any given utterance, the very same word in much the same context, is highly unstable, typically consisting of an almost random ordering of a few articulatory gestures (Menyuk & Menn, 1979). As a particularly compelling example of this, Ferguson and Farwell (1975) present the case of a 15-month old girl who produced 10 different variants of the target word /pen/ in a half-hour session, which they described as an attempt "to sort out features of nasality, bilabial closure, alveolar closure and voicelessness" (p. 423). The gestures, but not the timing, seem to have been grasped at this point. The third piece of evidence suggesting a more global representation derives from observations that production of any given segment is often heavily influenced by surrounding context, much more so than in adult speech, yielding a high incidence of consonant assimilation as in [dut] or [guk] for *duck*. (For reviews, refer to Ferguson, 1986; Menyuk & Menn, 1979; and Studdert-Kennedy, 1987).

As evidence that the lexical representations themselves are reorganized as phonological development progresses, Menyuk and Menn (1979) provide data showing that children retain vestiges of an earlier organization once they have learned a new pronunciation; thus representations for a given lexical item seem to involve multiple levels. This contrasts with previous models of child phonology which emphasized that a representation remains nearly constant from the initial acquisition of a word and affected segments are modified wholesale as "phonological rules" are changed. Whereas it is clear that some modifications are system wide, it seems that underlying representations may also undergo developmental change.

The evidence for continued reorganization throughout the preschool years comes from both production and perception. In one of the few direct tests of the hypothesis that basic phonological representations become less syllabic and more segmental over childhood, Nittrouer and colleagues have performed a pair of experiments examining the perception and production of the same contrast (s/sh) before two different vowels (i/u) in children at 3, 4, 5 and 7 years of age and in adult subjects. Their hypothesis was that a syllable-based phonology would result in even greater interdependence among component gestures than would a segmental phonology. In the perception study, consistent with this hypothesis, young children's identification of the initial fricative was more influenced by the syllable information carried by the following vowel segment than it was by the frequency information that is temporally constrained to the fricative; 7-year olds and adults relied more on the frequency information (Nittrouer & Studdert-Kennedy, 1987; Nittrouer et al., 1989). Corresponding effects were observed in an analysis of production: in an examination of the acoustic structure of syllables produced by these same subjects, it was observed that /s/ and /sh/ were more strongly differentiated with age before each of the vowels, while the extent to which the production of /s/ and /sh/ was affected by transition information extending across the syllable decreased with age. These results were interpreted to indicate that "children initially organize their speech gestures over a domain at least the size of the syllable and only gradually differentiate the syllable into patterns of gesture more closely aligned with perceived segmental components" (p. 120).

Other pairs of perception/production studies are consistent with the findings of Nittrouer et al. (1989). Zlatin and Koenigsknecht (1975, 1976) compared perception and production data on the use of voice onset time to discriminate minimal pairs such as *bees/peas*. In both sets of data, voice onset time measures reflected more discrete and less variable phoneme boundaries with increasing age. And, adult-level performance was achieved earlier for perception than for production. In a study of vowel duration as a cue to post vocalic

voicing (*bip* versus *bib*, *pot* versus *pod*) as evident in perception and production, Krause (1982a, 1982b) also reported that "both systems demonstrate refinement and stabilization with increasing age" (1982b, p. 25). In sum, three pairs of studies have directly addressed the hypothesis that phonological representations become increasingly segmental over this period; all have observed continued greater influence of syllabic structure on phonological contrasts in children up to eight years of age.

Further suggestion that phoneme representations are continuing to undergo change during the preschool years derives from an analysis of numerous spontaneous speech errors produced by two young girls studied between one and five years of age (Stemberger, 1989). This is a particularly interesting paradigm in light of the fact that speech errors have often been used as the most solid evidence that phonemes are, in fact, units of production and perception in adults (Shattuck-Hufnagel, 1987; Shattuck-Hufnagel & Klatt, 1979). Although Stemberger's children did produce errors involving the exchange of single phonemes (e.g., "you catch grap-hossers"), they produced proportionately fewer of these errors, and proportionately more feature errors involving subphonemic exchanges (e.g., "I got that gall for pristmas" for "ball for Christmas") than have been reported for adults. This result is consistent with the view that the phoneme may be a less cohesive unit in children than in adults. Similar findings emerged in an analysis of errors produced by children and adults in a word learning task devised by Treiman and Breaux (1982). Four-year olds were more likely to confuse labels which were "globally" similar (essentially where syllables would share features but not phonemes, e.g., *bis* and *diyz*), while adults tended to confuse lexical items sharing a common phoneme (e.g., *bis* and *boon*).

The apparent developmental shift in organization observed in production and perception shows interesting parallels with the development of phonological awareness over the same preschool and early school period. Highlighting this parallel is increasing evidence that many young children are sensitive to the sound structure of language, but only when the tasks involve units of organization larger than the phoneme, such as the syllable or syllable onset. Children as young as three years can categorize words on the basis of rhyme, alliteration and syllable structure; and yet investigators remain unable either to demonstrate or to instill awareness of phoneme segments prior to five years of age (Content et al., 1986; Fox & Routh, 1976; Liberman, et al., 1974; McLean et al., 1987; Walley et al., 1986). Treiman and Breaux (1982) examined this parallel with a direct comparison of phonological organization and awareness. Using the same stimuli employed in their word learning task, Treiman asked a different set of children and adults to classify which two words in a triad "go together." Consistent the results of the word learning study, four-year olds tended to group words on the basis of overall similarity, whereas adults were more likely to apply the criterion of a single shared phoneme. With training, adults could learn to classify on the basis of overall similarity, consistent with a multiple representation hypothesis. Children, however, could not be trained to classify on the basis of a single common phoneme.

To summarize the developmental findings, Nittrouer et al. (1989) proposed:

> that the initial domain of perceptuomotor organization is a meaningful unit of one or a few syllables. As the number and diversity of the words in a child's lexicon increase, words with similar acoustic and articulatory patterns begin to cluster. From these clusters there ultimately precipitate the coherent units of sound and gesture that we know as phonetic segments. Precipitation is probably a gradual process perhaps beginning as early as the second to third year of life when the child's lexicon has no more than 50-100 words. But the process is evidently still going on in at least some regions of the child's lexicon and phonological system as late as 7 years of age (p. 131).

III. IMPLICATIONS FOR READING DISABILITY

If phoneme-level representations are continuing to develop and to be refined over the preschool years, and if developmental changes in phonological awareness reflect changes in the very nature of lexical representations, what are the implications for understanding and explaining reading disability? I should note first, that such a finding would in no way alter our understanding of phoneme awareness as the immediate prerequisite to reading success. No matter how such awareness arises, it remains the case that reading an alphabetic script requires access to a phonemic organization of lexical structures. And neither would such a finding alter the long understood fact that awareness of this level of organization does not arise spontaneously in the normal course of acquiring a language.

What may be altered, however, is our understanding of how phonological development sets the stage for phoneme awareness, and why reading disability is associated with a broad array of phonological deficits. In regard to acquiring phonemic awareness, the effect of an increasingly segmental organization would seem to be straightforward. If, as the studies cited earlier suggest, the phoneme as a "crystallized perceptuomotor structure" (Studdert-Kennedy, 1987) begins its development in the second year of life and is still continuing up until at least seven years of age, one would not expect very young children to be able to segment syllables into phoneme-sized units, or to identify "common" phonemes across a range of syllable and word contexts. Rather, the cohesiveness of these phoneme structures in production, perception and memory may well be expected to influence the ease with which awareness of these structures is achieved, as is suggested by the general parallels that can be drawn between work on preschool phonological awareness and work on developing phonological representations. Similarly, children who, for whatever reason, are progressing more slowly with respect to these phonological abilities should experience greater difficulty in achieving phoneme awareness.

It is somewhat less obvious whether the hypothesized changes in phonological structures can explain the subtle deficits in phonological processing that have been observed in poor readers. Such a discussion requires one to consider what function a phoneme-level organization serves in memory, perception, production, and naming. One can, however, speculate that integrating gestures into phonetic routines serves as a mechanism for automation, providing a highly efficient representational code for encoding, storing and retrieving phonological structures in verbal working memory (Baddeley, 1986; Stemberger, 1989; Studdert-Kennedy, 1987). The same advantages of a phoneme-level representation invoked by Gleitman and Rozin (1977) to justify the utility of an alphabetic orthography should apply to underlying phonological representations as well. Consider the consequences of representing lexical items in terms of thousands of English syllables that vary in the identity and precise timing of individual gestures, and contrast that with a representation based upon approximately 40 phonemes whose gestural consequences have become well-specified and overlearned in the course of development. As argued by Gleitman and Rozin, a syllable-level representation may be ideal for Japanese which has 50 distinct syllables altogether, but is not at all efficient in English which has thousands of distinct syllable possibilities by virtue of cluster combinations and complex vowel systems and whose syllable boundaries are often obscure.

The effect of a segmental representation, then, may be to enable a child to convert (or "encode") the acoustic signal specifying a word into a sequence of discrete elements for storage and later reproduction of the correct articulatory shape. In contrast, a syllable-level or word-level encoding, in which a greater quantity of gestural information must be specified, may more readily overload the limited storage system, particularly in the case of phonologically complex items or lengthy strings of nonsense syllables. This would lead to a more variable and underspecified output, in which only the most salient features are recalled. It is interesting that in a lexical access task involving long complicated words like *thermometer*, poor readers could typically identify overall acoustic shape (e.g., producing *tornado* for *volcano;* or *bulb* and *gulb* for *globe*), but could not provide a full specification of the word (Katz, 1986).

Because the developmental model under discussion allows representations to become augmented over time, highly familiar lexical items would eventually become fully specified, whichever method (phonemic or gestural) is employed and we would be unlikely to observe difficulties specific to poor readers. Where one would expect to find differences to be more evident is in the case of novel words or pseudo-words where encoding strategies would play a larger role; or in the case of words which are phonologically complex, where specification may not yet be complete. Young children, and potentially poor readers, whose phonological representations are not segmentally organized may not be as efficient in assigning novel stimuli into a recoverable representation (i.e., for word repetition) as would be children who can readily assign a segmental structure. In fact, poor readers appear to be at a particular disadvantage when asked to repeat pseudoword stimuli (Snowling, 1981; Snowling et al., 1986), and multisyllabic words (Brady, Poggie, & Rapala, 1989). Furthermore, consistent with the developmental hypothesis under consideration, these same difficulties appear to be mirrored in younger preschool children presented with the same tasks (Brady et al., 1989).

It would seem that any task that requires reconstruction of the syllable would be aided by a segmental analysis and by refined, well-articulated prototypes of those segments to which

the input must be compared. This kind of reconstruction may be required when the signal is less than optimal either because of noise (Brady et al., 1983), artificial synthesis (Luce, Feustel, & Pisoni, 1983), or dialect variation (Mattingly, Studdert-Kennedy, & Magen, 1983). As discussed earlier, poor readers do appear to have special difficulty in reconstructing lexical items presented under noisy listening conditions (Brady et al., 1983). This finding, together with the fact that poor readers are also at a disadvantage when asked to recall a single nonsense-syllable after a filled interval (Dreyer, 1989), suggests that memory problems cannot be wholly explained in terms of lexical access or rehearsal strategies. Rather, both studies speak to a specific problem with encoding, in establishing the original structure of the representation.

In the developmental perspective presented, control over articulation is an important facet of developing stable, and ultimately accessible, phoneme categories (Studdert-Kennedy, 1987). Thus, one might expect to find an association between articulatory control and phoneme awareness. Poor readers should not be, and typically are not, characterized by an absolute inability produce one or another sound. Efficiency, rather than absolute ability, seems to be implicated in a study by Brady et al. (1989) in which poor readers and young children are more prone to error than good readers and older children when asked to repeat quickly sequences they can produce in isolation (e.g., *bu blu*). If articulation is an important prerequisite to acquiring phoneme structures, then explicit instruction in monitoring articulatory cues should aid in developing more crystallized units for structures to be segmented into. It is of some interest then, that the phoneme awareness training programs that seem to be particularly effective have involved an articulation component (Lindamood & Lindamood, 1969; see Lewkowicz, 1980, for a review).

IV. THE QUESTION OF CAUSALITY

If, as we suggest, individual differences in phonological awareness are in large part a function of differences in underlying phonological structures, then we have succeeded in moving the question only one step back. What then accounts for individual differences in the development of these structures? Do changes in phonological representation arise spontaneously in the normal course of language learning, independent of metalinguistic demands, driven by the demands of efficiently and adequately storing large numbers of vocabulary items in terms of the fewest sets of features? Or, are phonemes themselves achieved by linguistic analysis, necessarily predicated on language play and alphabetic instruction? That is, does the conscious analytic task of imposing segments on the speech stream for the purpose of reading and writing an alphabetic script lead and shape changes in underlying representations? The answer to this question has important implications for our understanding of reading disability.

Two accounts have been put forth. On the one hand, the *phonological deficit hypothesis* suggests that chronic reading failure results from inefficient phonological processes that impede both acquiring phoneme awareness and processing spoken and written language. On this account, phonological deficits in production, perception, memory and naming should both precede and predict reading failure and should be independent of both general intellectual factors and environmental factors. Such deficits, one might imagine, should hinder progress toward crystallized phonetic representations, and thereby provide the necessary link for a number of phonological problems associated with reading disability. In support of the phonological deficit hypothesis are prospective studies unconfounded with other variables (e.g., inclusion of readers in the initial sample), that suggest that individual differences in memory and lexical access may presage reading disability (e.g., Jorm, Share, MacLean, & Matthews, 1986). A compelling piece of evidence derives from a recently completed study by Scarborough (1990). In that study, reading disability at seven years of age was significantly related to performance on basic measures of language structure at $2\frac{1}{2}$ years of age (MLU, syntactic complexity, naming). Reading ability was not related mother's educational level or mother's reading ability, or to the child's receptive vocabulary at age two or age seven. Although the language performance of the would-be poor readers lagged significantly behind the progress of would-be good readers from comparable families, the poor readers' language skills were not sufficiently delayed as to be characterizable as "language delayed" on standardized measures. This finding serves to keep reading disability separate from, though on a continuum with, language disability.

Alternatively, the *orthographic hypothesis* suggests that phoneme awareness tasks, and potentially other phonological measures associated with reading failure, depend not on underlying phonological representations, but on orthographic representations derived as a function of reading experi-

ence (Ehri, 1989). Proponents of this view suggest that those phonological tasks on which poor readers fail may be handled more efficiently by reference to the orthographic representation. That is, some of the phonological deficits associated with reading failure may result directly from a lack of experience with the orthographic code, rather than from deficits in the phonological representations relied on in speaking and listening (Derwing, Nearey & Dow, 1986; Faber, in press; Tunmer, 1988). Even if we do not rely strictly on orthographic representations in all phoneme awareness and phonological processing tasks, there are many reasons to think that metalinguistic factors may play a role in developing a phonemic representation. In support of such a hypothesis, a recent study by McLean et al. (1986) found that individual differences in phoneme awareness in three- and four-year old children were strongly associated with the ability to recite nursery rhymes, independent of parental background.[2] What they proposed was that exposure to language play (epitomized here by nursery rhymes) enables the child to become aware of phonemic units. While their argument applies to phonological awareness, it may be possible to extend it to phonological representations as well, with language play aiding in the construction of the relevant units of phonology. Indeed, an important role for language play has been hypothesized by child phonologists in order to account for further refinements in word representations in toddlers beyond the 50-word stage (Ferguson & Macken, 1983; Jusczyk, 1986; Macken & Ferguson, 1983).

Although it may well be that metalinguistic experience in general, and orthographic experience in particular, may aid us in refining our phonological representations, these findings need not commit us to the view that phonemes are arbitrary or epiphenomenal in nature. The standard arguments continue to apply, including evidence from linguistic description, speech error analysis, and the ease with which we come to use an alphabet code, once awareness has been achieved (e.g., Gleitman & Rozin, 1977; Studdert-Kennedy, 1987). Recall, for example, that the speech errors produced by Stemberger's young children did include phoneme exchanges, although these occurred relatively less frequently than in adult production. One must also be able to account for the fact that many Japanese children not explicitly taught an alphabetic system eventually do "discover" phoneme categories (Mann, 1987). What is changed in this account is only that phoneme-level representations, implicit as well as explicit, may not come for free, but rather must emerge over time, in the course of lexical expansion, language play, and, potentially, orthographic experience.

Of course, if changes in representation, like the acquisition of phoneme awareness, can be aided by a catalytic prompt (most notably, reading instruction), then a fundamentally more optimistic story can be told. It suggests that successful training in phoneme awareness may have important repercussions throughout the phonological system. Although the evidence of considerable growth in segmental organization during the preschool years points to a developmental trend towards segments independent of reading experience, there is reason to believe that such a trend can be given a nudge. Note, for example, the remarkable findings of phonological coding in memory in deaf children who have been taught to read (Hanson, in press). Clearly these students did not derive segmental representations through canonical language development; it is possible, however, that the extensive articulation training provided the deaf may aid in forming these representations.

In sum, whereas a traditional view of phonological awareness sharply distinguishes between a phonological and environmental accounts of reading disability, the developmental perspective presented here can comfortably handle both: biological predispositions of the language system may be shaped by linguistic experience. Whether segmental awareness affects phonological representations, or vice versa, or whether there is a complex interplay between the two (as seems to be the case for phoneme awareness and reading), finding a close correspondence between the two may increase our understanding about reading disability.

SUMMARY

I have hypothesized in this paper that developmental changes in phonological representation may set the stage for acquiring phoneme awareness, and hence, for reading acquisition. I have further suggested that a failure along this same dimension may be responsible for the finding that poor readers are characterized by deficient phonological representations as evident in diminished short-term memory, inability to encode phonological structures under stressed conditions, and underspecified lexical representations.

A great deal of work remains both to track the development of phonological representation over the preschool and early school years, and to assess the impact of these changes on memory, analysis

and other tasks. Nonetheless, we are left with a sense that the phonological representations upon which analysis depends are not a pre-set immutable part of language. Indeed, as suggested here, it is possible they may be influenced by increased vocabulary, word play, phonological awareness, and literacy. Once phonological representation is taken out of the realm of the invariant, it becomes possible to do productive research on the relationship between,and development within, awareness and representation. It allows us to be both more specific about what a phonological deficit consists of and more optimistic about remediating reading disability.

I would like to conclude this discussion by applauding Treiman for her thorough and systematic chronicling of the emergence of phoneme awareness and for her efforts to relate awareness and underlying representations. Where many have indicated a lack of phoneme awareness, Treiman, like Liberman before her, has pressed on to ask the more positive question of what units of sound are salient for the young child. Treiman's further refinement of this developmental progression within the context of modern phonology sets the stage for a more explicit account of growth and change in phonological representation than has been available to date, enriching our understanding of phonological processing and phonological development.

REFERENCES

Baddeley, A. (1986). *Working memory*. Oxford University Press.

Blachman, B. (1984). Relationship of naming ability and language analysis skills to kindergarten and first-grade reading achievement. *Journal of Educational Psychology, 76*, 610-622.

Bradley, L., & Bryant, P. E. (1978). Deficits in auditory organisation as a possible cause of reading backwardness. *Nature, 271*, 746-747,

Bradley, L., & Bryant, P. E. (1983). Categorizing sounds and learning to read - a causal connection. *Nature, 301*, 419-421.

Brady, S. (1986). Short-term memory, phonological processing, and reading ability. *Annals of Dyslexia, 36*, 138-153.

Brady, S. (in press). The role of working memory in reading disability. In S. Brady & D. Shankweiler (Eds.), *Phonological processes in literacy: A tribute to Isabell Y. Liberman*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Brady, S., Mann, V., & Schmidt, R. (1987). Errors in short-term memory for good and poor readers. *Memory & Cognition, 15*, 444-453.

Brady, S., Poggie, E., & Rapala, M. (1989). Speech repetition abilities in children who differ in reading skills. *Language and Speech, 32*, 109–122.

Brady, S. A., Shankweiler, D., & Mann, V. A. (1983). Speech perception and memory coding in relation to reading ability. *Journal of Experimental Child Psychology, 35*, 345-367.

Browman, C. P., & Goldstein, L. (1986). Towards an articulatory phonology. *Phonology Yearbook, 3*, 219-252.

Browman, C. P., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology 6(2)* 201-251.

Catts, H. (1986). Speech production/phonological deficits in reading-disordered children. *Journal of Learning Disabilities, 19*, 504-508.

Chomsky, N. & Halle, M. (1968). *The sound pattern of English*. New York: Harper & Row.

Clements, G. N., & Keyser, S. J. (1983). *CV phonology: A generative theory of the syllable*. Cambridge, MA: MIT Press.

Content, A., Kolinsky, R., Morais, J., & Bertelson, P. (1986). Phonetic segmentation in pre-readers: Effect of corrective information. *Journal of Experimental Child Psychology, 421*, 47–72.

Denckla, M. B., & Rudel, R. G. (1976). Naming of object-drawings by dyslexic and other learning disabled children. *Brain and Language, 3*, 1-15.

Derwing, B. L., Nearey, T., & Dow, M. (1986). On the phoneme as the unit of the second articulation. *Phonology Yearbook, 3*, 45-69.

Dreyer, L. (1989). *The relationship of children's phonological memory to decoding and reading ability*. Unpublished doctoral dissertation, Columbia University.

Ehri, L. (1989) The development of spelling skill and its role in reading acquisition and reading disability. *Journal of Learning Disabilities, 22*, 356-365.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in early infancy. *Science, 171*, 304-306.

Faber, A. (in press). Phonemic segmentation as epiphenomenon: Evidence from the history of alphabetic writing. To appear in *Language and Literacy: Papers from the Symposium*. John Benjamins.

Ferguson, C. A. (1986). Discovering sound units and constructing sound systems: It's child's play. In J. S. Perkell & D. H. Klatt (Eds.) *Invariance and variability in speech processes* (pp. 36–51). Hillsdale, NJ: Erlbaum.

Ferguson, C. A., & Farwell, C. B. (1975). Words and sounds in early language acquisition. *Language, 51*, 419-439.

Ferguson, C. A., & Macken, M. A. (1983). The role of play in phonological development. In K. E. Nelson (Ed.), *Children's language* (Vol. 4). Hillsdale, NJ: Lawrence Erlbaum Associates.

Fowler, A. (1988). Grammaticality judgments and reading skill in grade 2. *Annals of Dyslexia, 38*, 73-84.

Fowler, A. (1990). Factors contributing to performance on phonological awareness tasks. *Haskins Laboratories Status Report on Speech Research, SR 103/104*, 137-152.

Fox, B., & Routh, D. (1976). Phonemic analysis and synthesis as word attack skills. *Journal of Educational Psychology, 68*, 70-74.

Fudge, E. (1987). Branching structure within the syllable. *Journal of Linguistics. 23*, 359-377.

Gleitman, L. R., & Rozin, P. (1977). The structure and acquisition of reading: Relation between orthography and the structured language. In A. S. Reber & D. L. Scarborough (Eds.), *Toward a psychology of reading* (pp. 1-53). Hillsdale, NJ: Lawrence Erlbaum Associates.

Godfrey, J., Syrdal-Laskey, A., Millay, K., & Knox, C. (1981). Performance of dyslexic children on speech perception tests. *Journal of Experimental Child Psychology, 32*, 401-424.

Gough, P. B., Walsh, M. A. (in press). Chinese, Phoenicians and the orthographic cipher of English. In S. Brady & D. Shankweiler (Eds.), *Phonological processes in literacy: A tribute to Isabell Y. Liberman*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Hanson, V. (in press). Phonological processing without sound. In S. Brady & D. Shankweiler (Eds.), *Phonological processes in literacy: A tribute to Isabell Y. Liberman*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Jorm, A. (1983). Specific reading retardation and working memory: A review. *British Journal of Psychology, 74*, 311-342.

Jorm, A. F., Share, D. L., Maclean, R. & Matthews, R. (1986). Cognitive factors at school entry predictive of specific reading retardation and general reading backwardness: A research note. *Journal of Child Psychology and Psychiatry, 27*, 45-55.

Jusczyk, P. (1986). Toward a model of the development of speech perception. In J. Perkell & D. Klatt (Eds.) *Invariance and variability in speech perception*, (pp. 1–33). Hillsdale, NJ: Lawrence Erlbaum Associates.

Katz, R. (1986). Phonological deficiencies in children with reading disability: Evidence from an object naming task. *Cognition, 22*, 225-257.

Katz, R., Shankweiler, D., & Liberman, I. (1981). Memory for item order and phonetic recoding in the beginning reader. *Journal of Experimental Child Psychology, 32*, 474-484.

Krause, S. (1982a). Vowel duration as a perceptual cue to post-vocalic consonant voicing in young children and adults. *Journal of the Acoustical Society of America, 71*, 990-995.

Krause, S. (1982b). Developmental uses of vowel duration as a cue to post-vocalic stop voicing. *Journal of Speech and Hearing Research, 25*, 388–393.

Kuhl, P. (1987). Perception of speech and sound in early infancy. In P. Salapatek & L. Cohen. (Eds.) *Handbook of infant perception* (Vol. 2, pp. 275-382). New York: Academic Press.

Lewkowicz, N. (1980). Phonemic awareness training: What to teach and how to teach it. *Journal of Educational Psychology, 72*, 686-700.

Liberman, A. M. (1989). Reading is hard just because listening is easy. In C. von Euler (Ed.), *Wenner-Gren International Symposium Series: Brain and Reading*. Basingstoke, England: Macmillan.

Liberman, I. Y. (1973). Segmentation of the spoken word and reading acquisition. *Bulletin of the Orton Society, 23*, 65-77.

Liberman, I. Y., & Shankweiler, D. (1985). Phonology and the problems of learning to read and write. *Remedial and Special Education, 6*, 8–17.

Liberman, I. Y., Shankweiler, D., Fischer, W. M., & Carter, B. (1974). Explicit syllable and phoneme segmentation in the young child. *Journal of Experimental Child Psychology, 18*, 201-212.

Liberman, I. Y., Shankweiler, D., & Liberman, A. M. (1989). The alphabetic principle and learning to read. In D. Shankweiler & I. Y. Liberman, (Eds.), *Phonology and Reading disability: Solving the reading puzzle*, IARLD Monograph Series, Ann Arbor, MI: University of Michigan Press.

Lindamood, C. H., & Lindamood, P. C. (1969). *The A.D.D. Program: Auditory Discrimination in Depth*. Boston: Teaching Resources Corporation.

Luce, P. A., Feustel, T. C., & Pisoni, D. B. (1983). Capacity demands in short-term memory for synthetic and natural speech. *Human Factors, 25*, 17-32.

Lundberg, I., Frost, J., & Petersen, O. (1988). Effects of an extensive program for stimulating phonological awareness in preschool children. *Reading Research Quarterly, 23*, 263-284.

Lundberg, I., Olofseon, A., & Wall, S. (1980). Reading and spelling skills in the first school years, predicted from phonemic awareness skills in kindergarten. *Scandinavian Journal of Psychology, 21*, 159-173.

Macken, M. & Ferguson, C. A. (1983). Cognitive aspects of phonological development: Model, evidence, and issues. In K. E. Nelson (Ed.)., *Children's language* (Vol. 4). Hillsdale, NJ: Lawrence Erlbaum Associates.

Mann, V. A. (1987). Phonological awareness and alphabetic literacy. *Cahiers de Psychologie Cognitive, 7*, 476-481.

Mann, V. A., Tobin, P., & Wilson, R. (1987). Measuring the causes and consequences of phonological awareness through the invented spellings of kindergarten children. *Merrill-Palmer Quarterly, 33*, 365–391.

Mattingly, I., Studdert-Kennedy, M., & Magen, H. (1983). Phonological short-term memory preserves phonetic detail. Paper presented at the Acoustical Society of America, Cincinnati, Ohio.

McLean, M., Bryant, P., & Bradley, L. (1987). Rhymes, nursery rhymes, and reading in early childhood. *Merrill-Palmer Quarterly, 33*, 266–281.

Menyuk, P., & Menn, L. (1979). Early strategies for the perception and production of words and sounds. In P. Fletcher & M. Garman (Eds.), *Language acquisition* (pp. 49-70). Cambridge: Cambridge University Press.

Morais, J., Cary, L., Alegria, J., & Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? *Cognition, 7*, 45-64.

Nittrouer, S., & Studdert-Kennedy, M. (1987). The role of coarticulatory effects in the perception of fricatives by children and adults. *Journal of Speech and Hearing Research, 30*, 319-329.

Nittrouer, S., Studdert-Kennedy, M., & McGowan, R. S. (1989). The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. *Journal of Speech and Hearing Research, 30*, 319-329.

Pallay, S. (1986). *Speech perception in dyslexic children.* Unpublished doctoral dissertation, The City University of New York.

Perfetti, C., Beck, I., Bell, I., & Hughes, C. (1987). Phonemic knowledge and learning to read are reciprocal: A longitudinal study of first grade children. *Merrill-Palmer Quarterly, 33*, 283-319.

Pratt, A., & Brady, S. (1988). Relationship of phonological awareness to reading disability in children and adults. *Journal of Educational Psychology, 80*, 319-323.

Rapala, M. R., & Brady, S. (1990). Reading ability and short term memory: The role of phonological processing. *Reading and Writing*.

Read, C., & Ruyter, L. (1985). Reading and spelling skills in adults of low literacy. *Remedial and Special Education, 6*, 43-52.

Rosner, J., & Simon, D. P. (1971). The auditory analysis test: An initial report. *Journal of Learning Disabilities, 4*, 384-392.

Rozin, P. (1975). The evolution of intelligence and access to the cognitive unconscious. In J. Sprague & A. N. Epstein (Eds.), *Progress in psychobiology and physiological psychology* (Vol. 6). New York: Academic Press.

Scarborough, H. (1990). Very early language deficits in dyslexic children. *Child Development, 61*, 1728-1743.

Shattuck-Hufnagel, S., & Klatt, D. (1979). The limited use of distinctive features and markedness in speech production: Evidence from speech errors. *Journal of Verbal Learning and Verbal Behavior, 18*, 41-55.

Shattuck-Hufnagel, S. (1987). The role of word-onset consonants in speech production planning: New evidence from speech error patterns. In E. Keller & M. Gopnik (Eds.), *Motor and sensory processes of language* (pp. 17-51). Hillsdale,NJ: Lawrence Erlbaum Associates.

Smith, N. V. (1973). *The acquisition of phonology: A case study*. Cambridge University Press: Cambridge.

Snowling, M. (1981). Phonemic deficits in developmental dyslexia. *Psychological Research, 43*, 219-234.

Snowling, M., Goulandris, N., Bowlby, M., & Howell, P. (1986). Segmentation and speech perception in relation to reading skill: A developmental analysis. *Journal of Experimental Child Psychology, 41*, 489-507.

Stanovich, K. E. (1985). Explaining the variance in reading in terms of psychological processes: What have we learned? *Annals of Dyslexia, 35*, 67-96.

Stanovich, K. E. (1988). The right and wrong places to look for the cognitive locus of reading disability. *Annals of Dyslexia, 38*, 175-190.

Stanovich, K., Cunningham, A., & Feeman, D. (1984). Intelligence, cognitive skills, and early reading progress. *Reading Research Quarterly, 19*, 120-139.

Stemberger, J. (1989). Speech errors in early language production. *Journal of Memory and Language, 28*, 164–188.

Studdert-Kennedy, M. (1986). Sources of variability in early speech development. In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and variability of speech processes*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Studdert-Kennedy, M. (1987). The phoneme as a perceptuomotor structure. In A. Allport, D. Mackay, W. Prinz, & E. Scheerer (Eds.), *Language perception and production*, (pp. 67–84), London: Academic Press.

Tartter, V. (1986). *Language processes*. New York: Holt, Rinehart & Winston.

Torgesen, J. (1978). Performance of reading disabled children on serial memory tasks: A review. *Reading Research Quarterly, 19*, 57-87.

Treiman, R. (1984). On the status of final consonant clusters in English syllables. *Journal of Verbal Learning and Verbal Behavior, 23*, 343-356.

Treiman, R. (1985). Onsets and rimes as units of spoken syllables: Evidence from children. *Journal of Experimental Child Psychology, 39*, 161-181.

Treiman, R. & Baron, J. (1981). Segmental analysis ability: Development and relation to reading ability. In G. E. MacKinnon & T. G. Waller (Eds.), *Reading research: Advances in theory and practice* (Vol. 3, pp. 159-198). New York: Academic Press.

Treiman, R., & Breaux, A. (1982). Common phoneme and overall stimulating relations among spoken syllables. Their use by children and adults. *Journal of Psycholinguistic Research, 11*, 569-597.

Treiman, R., & Danis, C. (1988). Short-term memory errors for spoken syllables are affected by linguistic structure of syllables. *Journal of Experimental Psychology, 14*, 145-152.

Treiman, R. A., & Zukowski, A. (in press). Levels of phonological awareness. In S. Brady & D. Shankweiler (Eds.), *Phonological processes in literacy: A tribute to Isabelle Y. Liberman*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Tunmer, W. (1988). Metalinguistic abilities and beginning reading. *Reading Research Quarterly, 23*, 134-158.

Wagner, R. K., & Torgesen, J. (1987). The nature of phonological processing in the acquisition of reading skills. *Psychological Bulletin, 101*, 192-212.

Walley, A., Smith,L., & Jusczyk, P. (1986). The role of phonemes and syllables in the perceived similarity of speech sounds for children. *Memory and Cognition, 14*, 220-229.

Werker, J., & Tees, R. (1987). Speech perception in severely disabled and average reading children. *Canadian Journal of Psychology, 41*, 48-61.

Wolf, M. (1986). Rapid alternating stimulus naming in the developmental dyslexias. *Brain and Language, 27*, 360-379.

Wolf, M., & Goodglass, H. (1986). Dyslexia, dysnomia and lexical retrieval: A longitudinal investigation. *Brain and Language, 28*, 154-168.

Zlatin, M. A., & Koenigsknecht, R. A. (1975). Development of the voicing contrast: Perception of stop consonants. *Journal of Speech and Hearing Research, 18*, 541-553.

Zlatin, M. A., & Koenigsknecht, R. (1976). Development of the voicing contrast: A comparison of voice onset time in stop perception and production. *Journal of Speech and Hearing Research, 19*, 78-92.

FOOTNOTES

*This paper was originally written as a comment on a chapter by R. Treiman and A. Zukowski. Both the chapter and this comment will appear in S. Brady & D. Shankweiler (Eds.), *Phonological processes in literacy: A tribute to Isabelle Y. Liberman*. Hillsdale, NJ: Lawrence Erlbaum Associates.

†Also Department of Human Development, Bryn Mawr College.

[1]Note that although I use feature and gesture interchangeably in my discussion of the literature, the idea of an articulatory feature extending throughout the syllable and existing independently of phonemes (Tartter, 1988) is different in critical respects from the distinctive feature described in Chomsky & Halle (1968). Whereas the traditional feature is a linguistic unit crucially tied to the phoneme, articulatory gestures in the Browman/Goldstein model have been defined independently of the phoneme level and explicitly refer to patterns of movement (of lip, jaw, tongue, velum and larynx), executed with "communicative intent" and embedded within the syllable. In turn, the syllable can be defined as a unit of perception and production without reference to the phoneme.

[2]Because the accuracy of a child's recitation of nursery rhymes was the measure of interest, the findings by McLean et al., 1986, could also be construed as supporting the phonetic deficit hypothesis. Children could be less accurate not because oflack of exposure but because of less developed linguistic ability.

Haskins Laboratories Status Report on Speech Research
1991, SR-105/106, 65-68

Modularity and the Effects of Experience*

Alvin M. Liberman and Ignatius G. Mattingly†

Experience is essential to the development of speech in the child, so we should ask of any psychological theory of speech how the effects of experience are to be accommodated. What does the theory have to say about how the child learns which phonetic distinctions are relevant in his native language and about how he adjusts to the phonetic capabilities of his own particular vocal tract as it changes in size and shape? As we show in this paper, the more and less conventional views of speech account for such effects of experience in categorically different ways. Our aim is not to weigh these different accounts against the evidence, but only to use them to illuminate an important difference between the theories from which they follow.

Sound localization and experience. Before taking up the theories about speech, we should be as explicit as we can about two different ways in which experience might affect percepts and their relations to everything else. For that purpose, we begin with the example of sound localization. We do this not because sound localization is simple, for it is, in fact, marvelously complex, but because it is well understood and lends itself readily to displaying the theoretical choices with which we are concerned. Consider, then, two conceivable accounts of sound localization, together with their different implications for the ways in which experience might affect development.

By any account, the information for localization takes the form of interaural disparities of time of arrival and intensity (Hafter, 1984). Complicating the processes by which such information is used is the fact that the disparities vary, not only as a function of source location, but also of frequency, aural acuity, distance between the ears, and orientation of the head. Further complications arise, of course, when several sources are present at once.

On one theory, the currently prevailing one, sound localization is managed by a neural mechanism narrowly and exclusively adapted to cope with the attendant complications and derive location of a source of sound from the information about disparity. We will refer to this mechanism as a module, using the term in the sense of Fodor (1983). That such a sound-localizing module exists has been shown most plainly in experiments on the barn owl (Knudsen, 1984; Konishi, Takahashi, Wagner, Sullivan, & Carr, 1988). In these experiments, investigators have found cells in the inferior colliculus that respond selectively to sounds according to their location in space, and thus form a map. Moreover, it has been possible to observe some of the processes by which this map is derived. But what is of particular importance for our purposes is that, though the information critical for location takes the form of disparities of time and intensity, with appropriate corrections for frequency, source coherence, and phase ambiguities, the positions of the neural responses are fixed by coordinates that are purely spatial; the proximal stimulus dimension of disparity is not represented. Information about this dimension is therefore not available outside the localization module, and the owl presumably does not perceive it. All this being so, we assume that the spatially organized neural responses correspond to the perceptual primitives.

But it is conceivable that, contrary to the neurobiological data and the most obvious facts about the perception of sound location by human beings, someone might nevertheless maintain that the disparities *were* perceived, and that the organism's knowledge of sound-source location was the result of a higher-level or cognitive computation based on these perceived disparities. Certain disparities would be associated with certain locations, or perhaps there might be some

This work was supported by NICHD Grant HD 01994 to Haskins Laboratories.

heuristic computation for getting locations from disparities. Let us now consider how experience might be expected to have an effect, depending on whether the modular or cognitive account is the more nearly correct.

The experience of interest is the change in the relation between the location of the source and the disparity cues as a consequence of changes, either natural or deliberately produced, in the acuity of the ears or the distance between them. The young barn-owl's ear may be plugged by the curious experimenter; the child's ears get farther apart as its head grows. In either case, there is a change in the amount of disparity for a given deviation in source location from the midline or a given amount of head rotation. Yet the owl somehow makes the appropriate adjustments in its sound-localizing behavior, and the child continues to localize sounds correctly as it grows (Knudsen, 1988).

On the first and generally accepted view of sound localization, the effects of experience must take place within the module. Indeed, exactly this has been found to be the case in the asymmetrically deafened barn owl, for Knudsen (1988) has shown that the neural map is itself recalibrated so as to maintain the veridical relation, obtained before the owl was deafened. Thus, it is the perceptual representation itself that is "corrected," not the cognitive connection between this representation and others. Apparently, the module adapts to the new environment at a precognitive level. In the case of the child, one can plausibly suppose that something of the same sort occurs when the disparities change as the head grows bigger.

On the other view of sound localization, the organism would have first to learn—by trial and error, logical inference, or instruction—that a certain seemingly arbitrary range of perceived disparities was, in fact, relevant to sound localization, and then, more specifically, how each disparity was to be interpreted as location. Such learning would, of course, have to take into account the complication that, for a fixed location, the disparities are different as a function of frequency and, even worse, that, at each frequency, the disparities change as the head gets bigger. Altogether, a formidable cognitive task. Of course, the task is no less formidable as it is done by the sound-localizing module. The difference is simply that the module is specifically and superbly adapted to its complex task, and carries it out without taxing in the least such cognitive capacities as the child might have.

Thus, the effects of experience that are cognitive contrast with the precognitive kind most obviously in the locus of the effects. In the precognitive kind, the effects are, as we saw, on the internal workings of the relevant module and thus on the perceptual representations themselves; in the cognitive variety, on the other hand, they would have to be in the connection between those representations and others. A further difference is that, while the precognitive calibrations of the module are highly selective in regard to the environmental conditions they respond to, cognitive learning is obviously quite promiscuous, being capable of forming connections to a wide variety of representations. Thus, an animal can be taught to make any of an indefinitely large number of responses whenever it perceives sound at a particular location. But this would in no way affect the localization module or the perceptual representations it produces; those would have changed only in response to environmental conditions that alter the relation between interaural disparities and the location of the sound source.

Speech and experience. Turning now to speech, we see that the conventional view is analogous to the view of sound localization that is incorrect. For it is most commonly supposed about speech that its perception is underlain by processes and primitives no different from those of nonspeech (Crowder & Morton, 1969; Kuhl, 1981; Miller, 1977; Oden & Massaro, 1978; Stevens, 1975). All sounds, whether they convey phonetic information or not, are supposed to excite the same specializations of the auditory system and evoke such standard auditory primitives as pitch, loudness, and timbre. The perceived difference between a stop consonant and a Morse code signal are only in the particular values that are assigned to each component of a common set of perceptual primitives. There are no specifically phonetic primitives.

Since, on this view, phonetic structures are not marked as a distinct class, the child must learn, obviously by some cognitive process, which of the indefinitely many percepts that belong to a common auditory mode are relevant to phonetic communication and which are not, and then, more specifically, which percepts are to be assigned to which phonetic categories. In this respect, learning to perceive speech would be, in principle, something like learning to perceive Morse code. In practice, it would be very much harder, of course, because, unlike the dots and dashes of Morse code, the sounds of speech bear a peculiarly complex relation to the phonetic structures they convey.

One might suppose that, in trying to come to grips with this relation, the child would be aided by the results of experimenting with the acoustic consequences of his own articulatory gestures. But here again the conventional view imposes a considerable cognitive burden, for it assumes that the primitives of the speech-production system are not specific to speech, but are rather common to a general action mode. Therefore, the child would have to discover about the phonetically unmarked movements of his articulators, just as he would about the unmarked auditory percepts, which ones were relevant to phonetic communication and what the more specific nature of their relevance might be. And, since the motor primitives would have nothing in common with the perceptual primitives, they would have to be linked, and establishing those links would necessarily be a highly cognitive process, depending, for the most part, on unrestrained trial and error.

So, on the conventional view of speech perception, development would have to take place at a stage beyond the primitives that any module produces. Also, of course, it would be relatively unconstrained in regard to the nature of the signals, processes, or events, that become connected; so, in this respect, too, learning to communicate with speech would be like learning Morse code.

But there is another view of speech, one that has implications more in accord with the most obvious facts of language development (Liberman & Mattingly, 1985, 1989). On this view, there is a phonetic module, a biologically coherent system specialized for the production and perception of phonetic structures. The primitives of this module, common to production and perception, are the articulatory gestures that serve as the building blocks of the phonological system. Thus, the phonetic module produces primitive representations that are specifically phonetic, hence categorically set apart from all others. There is, then, no need for a cognitive process that enables the child to learn to attribute communicative significance to some arbitrarily defined class of otherwise undistinguished representations. Moreover, as in the case of the sound-localization module, experience calibrates and recalibrates the perceptual representations by processes that are entirely internal to the module. It is by means of this calibration that the child adjusts to the subset of phonetic gestures appropriate to his own language and to the changing anatomy of his vocal tract. Such precognitive calibration acts on specifically phonetic primitives; the effect of experience is to guide that calibration, not to teach the child how to translate nonphonetic primitives into phonetic categories. A consequence is that the only experiences that count for the module are these that are relevant to the phonetic environment.

It is particularly appropriate that we consider this matter in a book that honors James Jenkins, because there is an experiment by Jenkins and his colleagues that provides relevant data (Miyawaki, Strange, Verbrugge, Liberman, & Jenkins, 1975). This experiment was designed to assess the effects of linguistic experience on phonetic perception, and to find the locus of the effect. Using synthetic approximations to the syllables [ra] and [la] that differed only in the extent and direction of the third-formant transition, Jenkins and his colleagues found, first, that native speakers of English reliably sorted the syllables properly and showed a pronounced peak in discrimination at a point on the acoustic continuum of third-formant transitions that corresponded to the English phonetic boundary. Speakers of Japanese, on the other hand, discriminated the syllables very poorly, and their discrimination functions showed no signs of a peak at the point that corresponded to the English boundary. It is important that the two groups differed, not just in their ability to attach phonetic labels appropriately, but in the functions that were generated when they tried simply to discriminate one stimulus from another on any basis whatsoever. For this indicates that the American listeners did not perceive these stimuli as the Japanese did, and then, by some cognitive process, apply the phonetic labels their language had taught them. Rather, the difference was in the precognitive, purely perceptual aspects of the process. And, obviously, the difference was a result of the differing linguistic experience of the groups, for, as is well known, the [r]-[l] distinction is not functional in Japanese. But Jenkins and his colleagues also undertook to find out just what it was that linguistic experience had affected. For that purpose they tested the ability of the American and Japanese listeners to discriminate the critical third-formant transition cue when, in isolation from the rest of the syllable, it did not sound like speech, but rather like a nonspeech 'bleat.' The result was that the two language groups discriminated the critical acoustic cue equally well. Thus, effect of linguistic experience was specifically on the phonetic system, not more generally on auditory perception.

The results of the experiment by Jenkins and his colleagues accord well with the view advanced in this paper. Relevant linguistic experience acts

on the internal workings of a phonetic module, with the result that the effect is on the representation itself, not on the way it becomes cognitively attached to phonetic labels or prototypes that exist at some further stage.

REFERENCES

Crowder, R. G., & Morton, J. (1969). Pre-categorical acoustic storage (PAS). *Perception and Psychophysics, 5*, 365-373.

Fodor, J. (1983). The modularity of mind. Cambridge, MA: MIT Press.

Hafter, E. T. (1984). Spatial hearing and the duplex theory: How viable is the model? In G. M. Edelman, W. E. Gall, & W. M. Cowan (Eds.), *Dynamic aspects of neocortical function*. New York: Wiley

Knudsen, E. I. (1988). Experience shapes sound localization and auditory unit properties during development in the barn owl. In G. M. Edelman, W. E. Gall, & M. W. Cowan (Eds.), *Auditory function: Neurobiological bases of hearing* (pp. 137-149). New York: Wiley.

Konishi, M., Takahashi, T. T., Wagner, H., Sullivan, W. E., & Carr, C. E. (1988). Neurophysiological and anatomical substrates of sound localization in the owl. In G. M. Edelman, W. E. Gall, & M. W. Cowan (Eds.), *Auditory function: Neurobiological bases of hearing* (pp. 137-149). New York: Wiley.

Kuhl, P. K. (1981). Discrimination of speech by nonhuman animals: Basic auditory sensitivities conducive to the perception of speech-sound categories. *Journal of the Acoustical Society of America, 70*, 340-349.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21*, 1-36.

Liberman, A. M., & Mattingly, I. G. (1989). A specialization for speech perception. *Science, 243*, 489-494.

Miller, J. D. (1977). Perception of speech sounds in animals: Evidence for speech processing by mammalian auditory mechanisms. In T. H. Bullock, (Ed.), *Recognition of complex acoustic signals* (Life Sciences Research Report 5, pp. 49-58). Berlin: Dahlem Konferenzen.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., & Jenkins, J. J. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception and Psychophysics, Vol. 18(5)*, 331-340.

Oden, G. C., & Massaro, D. W. (1978). Integration of featural information in speech perception. *Psychological Review*, 85, 172-191.

Stevens, K. N. (1975). The potential role of property detectors in the perception of consonants. In G. Fant & M. A. Tatham (Eds.), *Auditory analysis and perception of speech*. New York: Academic Press.

FOOTNOTES

*In R. R. Hoffman & D. S. Palermo (Eds.), *Cognition and the symbolic processes, Volume 3: Applied and ecological perspectives*. Hillsdale, NJ: Lawrence Erlbaum Associates (in press).

†Also University of Connecticut, Storrs.

Haskins Laboratories Status Report on Speech Research
1991, SR-105/106, 69-82

Modularity and Dissociations in Memory Systems*

Robert G. Crowder†

A cynical attitude toward progress in psychology is that we simply move back and forth between well-defined polarities, on pendulum swings, without really getting anywhere. Personally, I much prefer the model of a helix, in which we can recognize steady progress in one direction while not denying oscillations of perspective on certain other issues. One unmistakable trend that has been sweeping across the behavioral and cognitive sciences is the advancement of genetic explanations over environmental explanations. This is easy to find in such diverse fields as linguistics, intelligence, personality, and mental health, to say nothing of medicine. My grandparents were staunch believers in genetic causation, too, but I like to think we now have better reasons for our attitudes than they did.

A different trend back to earlier ideas is becoming evident in the reappearance of *faculty psychology* in cognition, generally, and in the analysis of memory, in particular. Publication of Fodor's *Modularity of Mind* (Fodor, 1983) only celebrates this newest cycle. The histories of such topics as localization of function in the brain and the interpretation of intelligence put the trend in perspective. The impulse to subdivide memory into isolated subsystems should be appreciated as a manifestation of Fodor's views, with modularity itself manifesting a historical rhythm that governs our approach to many of the great issues.

TWO SORTS OF MODULARITY IN MEMORY

In this first section of the paper, I will identify two very different interpretations of what *modularity* could mean in memory, one that may well be generally accepted as conventional wisdom and the other controversial. These correspond, I think, to Fodor's "horizontal" and "vertical" modularity, and to Jackendoff's (1987) "representation-based modules" and "fundamental principles," but I will call them, less abstractly, *coding modularity* and *process modularity*. In later sections, I take the dissociation of short-term storage (STS) and long-term storage (LTS) as a case in point, raising caution about concluding in favor of separate memory systems. I will show that belief in Hebbian consolidation is quite general in the animal field, just as the STS/LTS distinction has been among students of human memory. I conclude that the evidence is equally fragile in both cases. Finally, the status of empirical dissociations will be discussed in general, including the cases of recognition and recall and of declarative and procedural memory.

Coding modularity

Any attitude toward information processing must acknowledge that different kinds of information use different parts of the brain. Trivially, the auditory and visual systems engage distinct input pathways and cortical projections, as do the so-called minor senses. The specialization of the left hemisphere for language processing has been more tantalizing for the cognitive psychologist, perhaps only because it gives tangible reality to the concept of language.

As students of memory, many of us have blandly recognized this left-right specialization and yet, at the same time, we have held to a sort of Lashley position about learning and memory—that they depend on having *something* between the ears, but it is all-purpose machinery (some say oatmeal) to be found there. Such reliance on all-purpose equipment causes no one to lose sleep, provided learning and memory are themselves considered specialized functions of the mind. This assumption is dubious, however.

The growing identification of perception and information processing as being synonymous with

I greatly appreciate the comments of my colleagues Donald Broadbent, Fergus Craik, and Michael Watkins on an earlier version of this paper, whose preparation was also supported by Grant BNS 86 08344 from the National Science Foundation.

learning, or memorization (Craik & Lockhart, 1972; Crowder & Morton, 1969; Kolers, 1975) led many of us to change our attitudes in an important way: Once memory is regarded as a by-product of information processing, the implicit concept of all-purpose memory storage dissolves. To me, a comfortable language to describe this attitude is to say that *memory is not a storage process as such; it is simply the property of information processing that extends in time afterwards.*

In much the same way, the neutrinos detected at several international observatories starting on February 24, 1987, were not a sort of time capsule, laid down by Supernova 1987A for our benefit. These neutrinos represent part of that original event itself, in a galaxy called the Large Magellenic Cloud—an event that occurred 163,000 years ago and a billion billion miles away from here. These neutrinos *are* the event, observed at a remote context. And so it is with at least some kinds of memories. The retention is just an aspect of the original episode itself, manifest at some temporal remove. To me this attitude is the essence of proceduralism.

This attitude makes notions like *memory stores*, specialized to hold information over time, superfluous. If the memory is in some sense, an aspect of the original event, then the memory resides wherever in the nervous system the event did. The Proustian mechanism of redintegration based on olfactory/gustatory cues must then reside in the brain structures specialized for olfactory/gustatory processing, and of course the connections of these structures with others. Recognition memory for a kaleidoscope slide, for a snowflake, for a musical timbre, or perhaps for a spatial layout, would activate the participating units from their own respective information-processing episodes. Presumably, in these cases the processing, and therefore the memory residue, would have been scarcely verbalized. And when the information processing of interest consist, all along, of linguistic manipulations, the success or failure of retrieval efforts will depend on whether the originally active systems get reactivated by the cues at hand. I consider this view of episodic memory to be faithful to Tulving's (see Tulving, 1983) *encoding specificity principle*. It is also profoundly compatible with Hebb's (1949) ideas on cell assemblies and phase sequences as the agencies of memory. Related ideas in the animal-physiological approach to memory are found in Squire (1987, chapters 5 & 8).

Process modularity

So, coding modularity in memory is the inevitable consequence of (a) brain specialization in cognition and (b) proceduralism in memory. What Fodor calls *horizontal modularity*, which I call *processing modularity*, is quite another story. In this case, the separate subsystems cut across information formats (coding, or vertical modularity) in favor of common processes. An example that now seems rather crude, and which was cited derisively by Fodor (1983, pages 13-14), would be to propose distinct short- and long-term memory subsystems in which all sorts of memories would belong to one system up until some number of seconds had elapsed, and then to another system afterwards. Another case of process modularity is the distinction between procedural and declarative memory, increasingly popular in the 1980s. The same original experience, according to this last distinction, gets entered into distinct memory systems, one procedural, accessible to implicit memory tests, and intact in the amnesic, the other declarative, accessible to explicit probes of memory and seriously compromised in the amnesic.

What does it mean to propose distinct memory systems? Perhaps an analogy with visual information processing is instructive: It is now conventional to separate visual processing pathways, and their associated cortical relay sites, into two broad classes, some concerned with the identification of objects and others with their location in space (see, for example, Squire, 1987, p. 69). In addition, the same sort of distinction might cover sound localization and pitch perception, where the same stimuli might undergo processing in the two systems, for different purposes.

In the remainder of this paper, I shall try to examine the evidence and arguments that once made us embrace the distinction between short- and long-term storage. The case was grounded in solid empirical dissociations. I conclude, as I have before (Crowder, 1982) that the case for two memory systems is weak, even when the distinction is broadened to cover the Hebbian consolidation theory of memory.

On the basis of empirical dissociations between implicit and explicit memory tests, some are now urging us to make a distinction between another processing dichotomy, procedural and declarative memory (Squire, 1987, chapter 11; Tulving, 1987). These systems of memory would constitute processing modules of just the sort Fodor and

others have warned against. We can see in both the dual-trace approach to memory storage (STS/LTS), and also in the more recent procedural/declarative distinction, that empirical demonstrations of *dissociation* are central to the arguments. Why have the dissociations of the 1960s come on hard times? What does this tell us about the usefulness of empirical dissociations in proposing memory subsystems? I will argue that Tulving (1987) is too eager to make this inferential leap from processing dissociations to separate subsystems. Before going on, however, I should affirm that I attach the highest importance for theory to dissociations between and among memory variables. I take that importance to be beyond argument. The only question is whether they are sufficient to differentiate memory systems.

THE CASE OF SHORT-TERM STORAGE

I have reviewed elsewhere (Crowder, 1982) my revisionist interpretations of the concept of short-term storage (primary memory), and so I need not go into detail here. In a few words I will try to point out how the evidence relied on empirical dissociations. I pause to do that because the term *dissociation* was not popular in mainstream cognition until the exciting work on amnesics enriched our language with this particularly medical connotation.

Review of evidence

Recency. The traditional association of the recency effect in free recall with some transient memory has now been discredited by the work of Bjork and Whitten (1974; Whitten & Bjork, 1972, see also Baddeley & Hitch, 1977; Greene, 1986; Tzeng, 1973; Watkins & Peynircioglu, 1983). They showed that a post-list period of distraction, sufficient by itself to disrupt normal recency, does not eliminate recency if the items were spaced sufficiently from one another during presentation. If the distractor eliminates short-term memory then this long-term recency effect must not be caused by short-term storage. Maybe the "normal" recency effect is not caused by it either.

I fully appreciate that this inference is not universal: For example Schneider and Detweiler (1987) preferred to associate the short-term and long-term recency effects with different underlying mechanisms. Although I agree that parsimony is not a serious virtue in the theory of cognition, still, I think the burden of evidence must lie with those who wish to impose a complicated theoretical interpretation on a simple data pattern. If two creatures both look like goldfish and both act like goldfish, we should be prepared to accept that they are, in fact, *not* both goldfish, but the job of proving the distinction belongs to whomever believes in it.

Brown-Peterson forgetting. The second signature of the short-term storage system was the slope of the Brown-Peterson forgetting curve. Again, I have detailed elsewhere (Crowder, 1976) how this evidence was handled by the two-store theory and the argument is quite standard. Whether one's theory of loss in the Brown-Peterson task is an appeal to temporal decay, to displacement, or to perturbation of ordered rehearsal cycles, it *overpredicts* forgetting, specifically on the first trial of the experiment. That is, these mechanisms say nothing about proactive inhibition. There is nothing in these theories about why decay, displacement, and perturbation, respectively do not occur or have only negligible effects on the first trial of an experimental session. Far less do they anticipate that a change in the meaning category of the stimulus items would make decay, displacement, or perturbation suddenly become ineffective; nor why allowing a long intertrial interval would nullify these three mechanisms.

Broadly, then, a theory of Brown-Peterson forgetting has to be a theory of how proactive inhibition works in this task (see Crowder & Greene, 1987a). Theories of proactive inhibition tend to fall into two main categories: First, some assume that the essence of proactive inhibition is really negative transfer—subsequent items never get learned as well as the items presented on the very first trial. Besides the studies reviewed in Crowder (1982), a recent application of this second idea can be found in Schneider and Detweiler (1987) Second, some theories assume that proactive inhibition represents a problem of recency discrimination among traces that are widely available: We must always remember which stored items were those that appeared on only the most recent trial. This latter hypothesis is particularly well equipped to handle the findings listed above, as Gorfein (1987) has explained and as I shall assert again below. Notice that neither hypothesis for proactive inhibition has anything to do with special-purpose mechanisms of short-term storage. Both are well-understood principles of memory, plain and simple.

But recency and the Brown-Peterson slope were originally considered evidence for STS mainly because of *dissociations*. Recency recall and prerecency recall were sensitive to different

variables. Recency depended on distraction but recall from early and middle positions did not. The asymptote depended on such factors as word frequency, rate of presentation, and age of the subjects, but the recency effect did not

The analysis of Brown-Peterson forgetting depended on the *theoretical* dissociation first advanced by Waugh and Norman (1965) between the asymptote and slope of forgetting functions. But this assumption never had much in the way of experimental test because experimenters never had the patience to collect enough data points to allow separate estimates of these two performance parameters. This effort would have required curve-fitting based on tests of many different retention intervals within the same experiment.

Alternative account using temporal coding

Both the recency effect and the Brown-Peterson slope were dissociated from other performance measures in the same task (prerecency recall and the Brown-Peterson asymptote, respectively). There is a good case that both were originally misinterpreted. Furthermore, both results can now be explained in essentially the same way. They may depend on retrieval by means of the temporal context.

Recency has periodically been described in terms of "temporal" coding (e.g., Murdock, 1960). Recently, a temporal-discriminability approach to recency has been suggested by Glenberg (1987; Glenberg & Swanson, 1986).[1] The reasoning is that events that just happened are more distinctive with regard to their time of occurrence than are events from the more distant past. If time of occurrence were to be a retrieval cue, the recency advantage would result. Bjork and Whitten (1974) evoked a principle of "temporal perspective" in which an evenly spaced list of items was approached, for purposes of recall, from the recency direction, much as a line of evenly spaced telephone poles could be seen in spatial perspective from the near end. They assumed that the temporal discriminability of the recency items would be better than that of earlier items to the extent that the items were themselves spaced widely in time. Recall might depend on a sort of Weber-fraction, expressing (a) distance of the observer from the most recent event, as a proportion of (b) the spacing between events. The ratio would be increased either by recalling promptly after occurrence of the most recent item, or alternatively, by increasing the interitem spacing. The rule seems to predict recall well (Glenberg, Bradley, Kraus, & Benzaglia, 1983; Glenberg, Bradley, Stevenson, Kraus, Tkachuk, Gretz, Fish, & Turpin, 1980; Hitch, 1985). In a general sense, then, recency and temporal coding seem to have been joined in a coherent and quite general principle of human memory. Such a theory is able to address with one voice the long- and short-term recency effects, unlike the dual-trace position. As such, it must be taken seriously as a possible explanation of why people remember recent experiences better than distant ones.

It is not hard to imagine that factors such as word frequency, intelligence, presentation rate, and age would have minimal effects on temporal coding, and hence be ineffective in controlling recency. These variables would, however, continue to exert their familiar positive effects on degree of learning, and so affect performance where temporal coding has little impact, on the pre-recency segments of the serial position curve. Thus, the empirical dissociations on which the case for two stores was based are equally plausible from the alternative account.

One approach to the Brown-Peterson task, one that dispenses with talk of separate short-term memory systems, also stresses temporal coding. Bennett (1975) and Gorfein (1987) have advanced specific versions of this approach. They assume that subjects always face the problem of distinguishing the most recent memory items from those that occurred on earlier trials. On the first trial, these most recent memories are the only ones eligible and so the "discrimination problem" disappears and there is no forgetting. After the first trial, the difficulty in making the recency discrimination depends on several factors. One is the retention interval: Even if there are many competing items from earlier trials, if the most recent one has just been presented, with no interpolated material, it is still in the foreground of the temporal context. Lengthening the retention interval destroys the special foreground privilege of the most recent item, perhaps in conformity with Glenberg's (1987) "ratio rule" for recency in general. This accounts qualitatively for why there is no forgetting on the first trial of a Brown-Peterson experiment and for why forgetting increases with retention interval after the first trial. Gorfein's position also easily accounts for data on shifts in taxonomic category of the memory items and data showing sensitivity of Brown-Peterson forgetting to the interval elapsing between trials.

I cannot guarantee that this sort of approach will carry the day, in the end, for the Brown-Peterson task. But the account is very good now,

and so the burden of evidence must now fall on those who believe in some mechanism of short-term storage, such as a limited-capacity buffer, decay, or the perturbation principle. The detailed facts of forgetting in Brown-Peterson experiments do not demand any such extra principle, so these advocates must find other sorts of evidence to support them. Furthermore, these principles all make one crucial prediction for which evidence is weak: They expect that there should be memory loss on the very first trial of a Brown-Peterson experiment. Nothing in perturbation, limited capacity, or buffer principles says that "all bets are off" during the first experimental trial. Indeed, our lives are full of "first trial" situations, with the varieties of experiences outside, as opposed to inside, the laboratory. So, the first trial is not an exceptional circumstance, to be set aside as a fluke; it is ecologically central to our task as theorists. Let us examine first-trial forgetting for a moment:

Forgetting without PI? A survey of the literature

A crucial question for theory is therefore whether, after all, people forget on trial one of a Brown-Peterson experiment. I have been amazed that this question is not stressed more given the quality and quantity of evidence available and given the importance for theory of the question. Baddeley and Scott (1971) deliberately sought good evidence on the point, explicitly appreciating its importance. (Part of the problem with investigating this issue is that a subject is wasted after his or her first few trials in an experiment, a matter of a few minutes, and must then be sent home.) Baddeley and Scott combined individual studies by Baddeley and his associates in which a single trial had been given, in order to amass sufficient observations to trace first-trial forgetting curves in detail. The total *n* was an impressive 922. The combined result showed a reliable decline over the first 5 seconds, but the function was not monotonic, showing improvement in performance between 6 and 9 seconds. So this ambitious effort leaves some readers unsatisfied and in need of additional evidence.

Most of the many remaining experiments published on proactive inhibition set a fixed retention interval, often 20 seconds, and examined build-up and release as a function of other variables. As Baddeley and Scott (1971) said, of the data permitting estimation of trial-one performance, some are furthermore uninterpretable owing to ceiling effects on the first-trial—notably the famous and frequently reprinted data of Keppel and Underwood (1962) themselves.

Among the few adequate studies, Loess (1964) showed no loss on the first trial in two independent experiments. In the first, performance was near the ceiling at some intervals (scores of .958, .875, and .958, respectively, for intervals of 9, 18, and 27 seconds. In Loess's second experiment the first-trial scores were .79, .75, and .79 for intervals of 3, 9, and 18 seconds. In both experiments there were 24 observations per data point. Noyd (1965; see Fuchs & Melton, 1974) tested independent groups of 27 subjects on either two-, three-, or five-word items following delays of either 4, 8, or 24 s of digit reading. The two-word items were at ceiling. Performance on the three-word items was .927 at each interval, perhaps also uncomfortably close to the performance limit. Corresponding results for the five-word items were .676, .630, and .638, respectively. The decline between 4 and 8 seconds with the five-word items corresponds to 23% of a word and was not reliable, and so in this experiment there was no loss over time in the absence of proactive inhibition. Cofer and Davidson (1968) tested 18 subjects each at 3 or 18 seconds of counting backwards on three-consonant syllables and obtained perfect recall rates of .78 and .83, respectively. This gain as a function of retention interval was nonsignificant. Wright (1967) tested 240 people, 80 each at 3, 9, and 18 seconds of counting backwards and obtained correct recall proportions of .93, .90, and .96, respectively. These proportions are close to the performance ceiling, but the number of observations per data point is commanding and what trend might be evident in the data is not a declining one. Turvey, Brick, and Osborn (1970) tested subgroups of 40 subjects on three-consonant items at 5, 10, 15, 20, or 25sec, obtaining recall probabilities of .87, .85, .93, .93, and .95. Finally, an experiment by Gorfein and Viviani (1980, data given in Gorfein, 1987) showed off-ceiling performance that did not deteriorate with retention interval.

So, what should we conclude about forgetting on the first trial of a Brown-Peterson experiment? Perhaps the most conservative conclusions is that there are some inconsistencies in the literature but that first-trial forgetting is the exception rather than the rule. At best, the role of any additional short-term storage or primary memory mechanism, over and above the principles responsible for proactive inhibition, is *empirically*

slight in relation to observed performance losses in the Brown-Peterson task. This task can therefore no longer be exhibited as the showcase for short-term storage. Only with faith and enormous effort can evidence for short-term storage be coaxed from this task.

New sources of evidence?

So much for the techniques once thought to bring short-term storage into the laboratory, recency and Brown-Peterson forgetting slopes. What new kinds of task are cited for the concept nowadays? In a recent review of the evidence for buffer storage, Schneider and Detweiler (1987) cited a third phenomenon, the *span effect*. The reference is to the fact that immediate ordered recall is limited to about a handful or two of items, depending in well-understood ways on what the memory items are. But the early hypothesis that this performance represents a fundamental constant in cognitive capacity (e.g., Miller, 1956) has not aged well. Watkins (1977) has demonstrated different forms of coding for the early and late serial positions of an eight-item memory-span list. Others have dissociated age effects on the primacy and recency segments (Cohen & Sandberg, 1977; Huttenlocher & Burle, 1976; Samuel, 1978). A more plausible candidate for memory span may be that achieved with running-memory-span tests (Crowder, 1969; Pollack, Johnson, & Knaff, 1959) but its size is more on the order of 2 or 3 items than 7 ± 2 (Glanzer, 1972). Such a miniscule memory span is unattractive to those who would like to give short-term storage a role in other cognitive tasks. As we shall see, the working memory system of Baddeley and Hitch (1974; Baddeley, 1983) does have informative things to say about memory span, but more that it results from a highly specialized trick—the articulatory loop—than that it is a fundamental manifestation of an all-purpose capacity to buffer information.

To my knowledge, new candidates for *epitomizing* short-term memory within an explicit testing format have not received widespread acceptance, the way the recency and distractor techniques were accepted in the past. Rather, the newer orientation is to see short-term retention as a participant in integrated cognitive functioning, as *working memory*, to which attitude we now turn.

Working memory

One commendable development in the academic study of memory has been its integration with ongoing cognitive tasks that are not, in themselves, memory tests. This began with the early computer models of associative memory (Anderson's FRAN, etc.) and has continued, for example, with models of reading (Daneman & Carpenter, 1983) and reasoning (Case, Kurland, & Goldberg, 1982). The introduction of the term working memory by Baddeley & Hitch (1974; see Baddeley, 1986) has underlined this increased ecological perspective on memory. From relatively simple beginnings the model of working memory has been elaborated steadily, and now stands as our most comprehensive theory of short-term memory (Baddeley, 1983, 1986). The evidence from experiments that functionally distinguish among components such as the articulatory loop and visuo-spatial scratchpad is convincing. Notice that these distinctions are largely based on coding differences rather than storage-time differences; they exemplify coding modularity and not process modularity.[2]

Working memory does not, for me, merely reorient the older concept of STS in a new dressing. In its most fully articulated components, it is more like a *bag of tricks*, each a modular coding format in the sense discussed earlier in this paper. The tight connection between the articulatory loop and specific motor codes (Baddeley, Thomson, & Buchanan, 1975; Ellis & Hennelly, 1980) illustrates this point admirably. So does the experiment of Reisberg, Rappaport, and O'Shaughnessy (1984) in which people were taught to use a simple, thoughtless, finger-tapping "loop" to hold an additional item or two for memory span, beyond what they could otherwise handle. This form of "buffer storage" is fundamentally procedural, highly code specific, and not at all representative of a "memory system" to be distinguished from LTS.

Necessity of buffer memory for cognition?

The utility of talking about working memory is that it stresses that people require—without any possible argument—to remember small packages of information briefly in order to succeed in any complex thinking task. But is this an argument for the necessity of memory per se in human cognition, or is it rather an argument for a separate memory subsystem that is different in structure or function from "regular" memory, whatever that is? Here, I think we drift off from issues on which evidence has been brought to bear. *Evidence for (a distinct subsystem of) short-term storage is not at all the same as evidence that people need to store things over the short term.* No quarrel is possible with the assertion that people need memory for the recent past in order to

engage in language comprehension, written or spoken, in problem solving, in musical cognition, in playing bridge, or in almost any reasoning or language activity. Perhaps the truest sense in which we need a working memory mechanism is that we need a memory system that works. Distinguishing theoretically the operating principles of (a) memory for recent events from (b) memory for remote events is a much stiffer assignment, and one that I find relatively neglected these days.

Why, then, does it seem so intuitively *correct* that we attribute working or immediate memory phenomena to a different system than passive, long-term memories? Part of the answer is that some proposed subsystems, such as the articulatory loop, do indeed have integrity as separate forms of information processing. Because these codes are especially useful in the short term, we are fooled into thinking they are distinctive because of their short-term properties, rather than because of their coding format.

The assumption of process modularity—horizontally distinct memory subsystems—must make a much stronger claim, namely that short- and long-term storage are different *within a common coding format*. Otherwise, if coding format and short- versus long-term status are confounded, there may be no need for the latter distinction. What evidence is there that a phonetic memory code used in digit span is different in kind than a phonetic memory code used in word encoding experiments like those of Fisher and Craik (1977), where words were cued by phonetic hints some 5 minutes after learning? I believe there is none.

Historically, the theoretical confounding of coding and process modularity probably arose in the following way: The early workers stressed mainly the coding distinctions in the belief that they were subdividing short-term storage. In other words, the lesson of coding diversity was first appreciated with short-term storage. As evidence accumulated that traditional memory—long-term storage—was comparably subdivided into coding modules, the implicit responsibility arose for distinguishing which type of modularity—coding or process—was the more important. That is the central issue faced in this essay, and the answer I give is obviously that coding modularity is well supported but process modularity is not.

In answer to our question of why primary memory seems so distinctive, I can only appeal to language so often quoted from William James (1890) on the nature of primary memory. With virtually unchanged internal and external context, information from the immediate past *seems* still to belong to the psychological present, in the sense of Tulving's (1983) "recollective experience." Memory and forgetting do not demonstrably obey different principles provided we equate for their coding format. That is, what marks *semantic coding in memory* is the same whether testing occurs without appreciable delay (which may be rare) or quite a bit later. What marks phonetic coding is likewise continuous between short and long testing delays, the same for olfactory coding and for visual-imagery. These different forms of coding may be more or less durable, perhaps because of the density of interference that occurs after learning, but different laws do not suddenly come into play with long retention intervals. The difference between long and short intervals is, in all cases, that the immediate test occurs with little contextual change and therefore the system of time perception registers almost no change, whereas at longer intervals the change has been considerable.

Thus, the same dissociations that led us to distinguish STS and LTS need not imply two separate memory stores, as we once thought. Coding (vertical) modularity proves to be the more valuable principle than process (horizontal) modularity, in memory theory as Fodor (1983) claimed for cognition in general. In the next section, I examine a closely related aspect of memory theory, consolidation, in order to trace parallels with the dual-store ideas.

Where did STS come from? The case of consolidation

I recently had occasion to review some work in the neuropsychogical theory of memory. I was reminded that we cognitive psychologists should not be possessive of STS, as if we had invented it, via such workers as Broadbent (1958), Brown (1958), and Peterson and Peterson (1959). Neuropsychologists believe that the concept is legitimately theirs, and indeed their case is a good one: The history of thinking on *consolidation theory* over recent decades is instructive when we consider dissociations and the evidence for a separable state of short-term storage. To my surprise, I found the same bankruptcy in the original concept of consolidation as I have in the concept of STS. As we shall see, a new conception of consolidation has emerged.

Hebb and Gerard

Hebb's (1949; see also Gerard, 1949) neuropsychological theory was the landmark in

the modern history of consolidation theory. His statement had a wide influence in what we now call the neurosciences, as well as its influence in psychology, anticipating as it did the popular dual-trace (STS/LTS) distinction. According to Hebb's version of the dual-trace hypothesis, experience is first recorded in the form of labile, reverberating, organized patterns of firing among neural units, which, if allowed to remain active long enough in concert, lead to the formation of structural changes in the nervous system, the basis of long-term memory. Notice how congenial Hebb's formulation is with the famous lines of William James that characterized primary memory as the persistence of (active) attention and secondary memory as memory proper. We see Hebb's ideas honored even more in the modern two-process theory of Estes (1972; Lee & Estes, 1981), about reverberatory cycles of ordered information giving way, with rehearsal, to structural representations of serial order.

The continued debate on the idea of memory consolidation has centered on several research areas. Of these, the two most prominent are (a) animal memory and (b) human clinical amnesia. As Weingartner (1984) said, the concept of memory consolidation, central as it was to many developments in the modern psychobiology of learning and memory "...was either ignored or rejected by investigators of cognitive processes in unimpaired human subjects" (p. 204).

Exceptions to Weingartner's observation are few: Interest continues sporadically in *sleep* as a factor in retention (Ekstrand, 1972). No doubt seems to exist that retention is better following an interval containing sleep than following one that does not (see also Hockey, Davies, & Gray, 1972). The theoretical weight of this fact is not easily measured, though. Subjects in conditions calling for sleep almost immediately after learning must be irresistibly drawn to rehearsal while they are "drifting off" to sleep. But Ekstrand (1972) cited evidence that memory is a reliable function of whether or not, during sleep, there has been rapid eye movement (REM) activity. REM activity indicates the presence of dreaming during deep sleep. One outcome reported by Ekstrand is that memory performance is better following periods without REM activity than following REM episodes. But other reports (Empson & Clarke, 1970) have selectively deprived people of REM sleep with resulting *damage* to recall performance (see Jones, 1979, for discussion and more citations on this point.) Comparing sleep with and without REM activity is obviously better than comparing sleep versus wakefulness, but the content of REM dreaming itself might provide interference (or sources of reinstatement of the learning activity).

Another use of the consolidation idea in research on normal human subjects is due to Landauer (1974; 1977). Essentially, Landauer's experiments show that a given amount of high-similarity interference, in short-term paired-associate learning, has a larger effect if it comes right after acquisition than if it comes after a delay. Landauer's (1974, 1977) results are consistent with consolidation theory in that perseveratory activity specifically promoting consolidation of an item is more likely to be broken apart by highly similar items than by an items sharing little with the original learning. Relatively speaking only, low-similarity interference may be said to correspond to sleep. If so the earlier perseveratory activity is disrupted, the more its consolidation for the long run should be compromised. But that result is also consistent with other ideas about memory and so it does not uniquely favor consolidation theory. For example, both the acid-bath theory of Posner and Konick (1966) and Estes' stimulus perturbation model (Estes, 1972) predict just this result. At any rate, we do not need to resort to perseveration-consolidation to explain the especially damaging effects of prompt, as opposed to delayed, interference. Now, the alternative explanations for this pattern of results may eventually boil down to formal equivalence with consolidation theory once consolidation theory is worked out in detail. For the moment, the evidence about consolidation from conventional experimentation is not nearly so powerful as the clinical evidence.

Wickelgren (1977, 1979) is the leading advocate of consolidation theory who is not *primarily* identified with research on amnesia. Wickelgren's treatment of consolidation (1977) distinguishes between two possibilities for what changes when perseveration is allowed to run its course. An hypothesis based on *unitary strength* would assume that the memory trace just gets stronger and stronger following learning. If this were true, reminiscence would be the rule rather than a delicate and elusive phenomenon. Alternatively, (a) the beneficial influence of consolidation following learning and (b) the detrimental influence of decay could be rationalized with some version of the STS/LTS distinction. However, it would be preposterous to assign the STS mechanism a role lasting up to several *decades*, and the data on ECT do indeed provide evidence that the consolidatory process extends over decades (Squire, Slater, & Chace, 1975).

Wickelgren's own formulation of consolidation theory, based on these rational considerations and on the study of amnesia, is called the *decreased fragility hypothesis*. This idea is a single-trace-strength hypothesis about memory representation, but it has a two-factor account of trace dynamics over time. As traces age, they undergo decay. But at the same time, they "grow in resistance to decay;" they decrease in fragility. Decay, for Wickelgren, need not be simple disuse with the passage of time. Resistance to decay (decreased fragility) is assumed to grow indefinitely as the memory trace gets older and older, but with diminishing returns, so that the first moments after learning are the most important ones.

Two historical "laws" of memory anticipate Wickelgren's notion of trace fragility. Ribot (1881) deserves priority in this: His *law of regression* states that the vulnerability of memories to disruption lessens with their age. Ribot derived this generalization from a survey of amnesia cases produced by head injury. As now, the evidence for this important proposition then came from the clinic more than from the laboratory. Jost's second law (1897; see McGeoch, 1942; Woodworth, 1938) says that *if two associations are of equal strength but different age, the older one diminishes less with time*. Jost had been interested in a related idea as manifested in experimental studies of distribution of practice.

Wickelgren showed (see summary in Wickelgren, 1977) that the detailed analysis of forgetting curves was consistent with a mathematical model including separate trace strength and trace fragility expressions. However, the main evidence was clinical, from amnesia cases, including the Squire, Slater, & Chace (1975) study of ECT. This study, and data on head injury seem to show that recent memory is disrupted according to a temporal gradient by a traumatic event. The "lost memories" are very definitely "still there," however, because (a) *before* an ECT session, subjects are fine at remembering material from the most recent decades (which is unavailable just after ECT), and (b) as Russell and Nathan (1946) pointed out for head injury cases, the memories spontaneously recover with the passage of time since the injury (that is, the amnesic material comes back in reverse order to its age).

The modern version of consolidation theory takes these facts seriously, and the rest of us should quit ignoring them. That is one lesson of this section. The other lesson is that Hebbian consolidation, the dual-trace theory, which was so congenial to STS/LTS distinctions, now has little to recommend it as a general theory, for the same two reasons just examined, as well as others. And what of research evidence from animals?

Evidence from animal experiments. One of the most persuasive reasons for experimentation on animals, rather than humans, is that radical manipulations comparable to the trauma of head injury can be produced at will with animals but not, ethically, with humans. Accordingly, the truly experimental analysis of consolidation has belonged to the animal laboratory for many years. With animals, the investigator is free to introduce some electrical or chemical agency that might block organized perseveration in the brain and the resulting consolidation. The time course of consolidation can then be studied by manipulating the delay between learning and the administration of such an amnesic agency.

Skipping over much history, we can focus on the experiment in which rats are taught to step down from a platform to escape shock to their feet. In experimental conditions, electroconvulsive shock (ECS) is administered at various delays after a single step-down trial. The question is whether it impairs storage of the remembered footshock. The results of the Chorover and Schiller (1965) experiment and other subsequent ones did indeed show amnesic effects of an ECS treatment following one-trial punishment training, but only when the time between initial acquisition and ECS was less than 10 seconds. A temporal gradient occurred in that ECS delays of more than 10 seconds gave results like control conditions without ECS. Hilgard and Bower (1975) have written a thorough review of subsequent developments in the animal ECS experiments. For now, the important point is that an experimental amnesia, with a temporal gradient, can be produced in rats. This observation supports the Hebbian consolidation theory of memory but even this support has not been unequivocal in light of further reports:

1. Experiments on different species (for example, mice) or even different strains of rat, or experiments using slightly different task details have turned in wildly variable time constants for consolidation, even using the Chorover-Schiller experimental logic (Chorover, 1976; McGaugh & Gold, 1974). At the least, consolidation time estimates frustrate those who would have hoped for a single "magic number" for such a *fundamental* brain process as consolidation.

2. Memories impaired by ECS just after learning can be recovered spontaneously just by a

lapse in time before testing, as Russell and Nathan, (1946) documented long ago for head injury cases in humans. Miller and Springer (1973) reviewed some of the evidence for recovery in the animal studies. If the temporal gradient of lost memories shrinks with the passage of time before testing, we cannot say for sure that materials "forgotten" in a test are necessarily unavailable. It is always possible that waiting a little longer would show recovery.

3. The effects of ECS can be reversed, or largely reversed, if a "reminder" is given during the interval before testing. In one situation, rats were given a simple footshock outside the training and test apparatus. This apparently (R. Miller & Springer, 1973) reinstateds the training contingency between stepping off the platform and footshock. If the memory can be reinstated by such a reminder, it must have been laid down after all, and not obliterated by disruption of the consolidation process.

For these and other reasons, some workers have chosen to regard the induction of retrograde amnesia in animals, by ECS, as compromising primarily the *retrieval process* and not the consolidation process as originally thought (Gold & McGaugh, 1984; McGaugh & Gold, 1974; Miller & Marlin, 1984; Zeckmeister & Nyberg, 1982). R. Miller and Marlin (1984) are especially emphatic in rejecting the consolidation interpretation of amnesia as it was originally intended by Müller and Pilzecker, suggesting that: "To define all retrograde disruption of acquired information as consolidation failure and then cite retrograde disruption of acquired information as evidence of consolidation failure is circular and does not add to knowledge..." (pp. 86-87).

Hebbian theory as an article of faith

However, these same authors (R. Miller & Marlin, 1984) endorse Hebbian memory consolidation almost on logical grounds: The very first moments following an experience must, they say, carry that experience in some form of activity trace. Nobody would maintain that a structural brain change occurs instantaneously. But a structural brain change must happen sometime, for it is unreasonable to assume all memory is an activity pattern even years later. Therefore, they said, there must be a transition between the two forms of storage, as Hebb proposed. Miller and Marlin report some analyses that suggest such consolidation can take place as rapidly as within 500 milliseconds following an experience.[3] Once this reasoning is accepted, then so is the consolidation hypothesis of memory. Whether consolidation is a concept that explains any *observable behavior* is altogether another question, however. If the reversibility of ETC amnesia and the recovery of memories lost to retrograde amnesia argue against disruption of consolidation in humans, now we find that little or no evidence can be cited for this hypothesis in animals, either.

Consolidation reinterpreted

Renewed interest in consolidation, now reinterpreted in terms of Ribot's Law, can determine whether this is just a different way of talking about familiar mechanisms (rehearsal, test events, and so on) or whether we have been missing out on important discoveries. More to the point for present purposes, the status of consolidation, as that idea was originally understood, turns out to be quite like that of short-term storage: Workers seem disposed to trust the idea, almost on sheer faith, without clear evidence that can be cited as uniquely favoring it.

Taking stock of the arguments presented here, we note that the discussion of consolidation theory has pertained most directly to the bankruptcy of the concept of short-term storage. Beyond this, one may well wonder whether I am not proposing to substitute one form of processing modularity for another, by revising what interpretation of consolidation is tenable in light of the evidence. This question, in turn, depends on whether consolidatory changes could one day be understood in terms of coding modularity. So far as I know, the issue has not been addressed. With agencies such as ECT and memories of human beings for information from the recent and remote past, changes in coding might well be important. With lower species, such reasoning appears fanciful. Currently, we must restrict ourselves to the more conservative point that the theory of consolidation does nothing to limit the generality of my earlier conclusions about short-term storage.

DISSOCIATING RECALL AND RECOGNITION

It would have been a mistake, I believe I have shown, to accept dissociations as grounds for distinguishing short- and long-term storage systems. Let us take still another example, briefly: If empirical dissociation were the criterion for differentiating memory systems, our field of memory might soon become a taxonomic science resembling botany. Then surely not only STS and LTS would have remained processing modules, but also recall and recognition would have. I have

organized the dissociating evidence for recall and recognition elsewhere (Crowder, 1976), but the best known factors are word frequency, intentional versus incidental learning, and semantic organization. The effects of these variables on recall are either opposite to their effects in recognition, a pattern called *crossed double dissociation* by Dunn and Kirsner (1987), or there are null effects in the case of one measure and positive effects in the case of the other (*simple dissociation*).

The recall/recognition dissociations are important theoretically. They serve to falsify single-process theories like strength, which might claim that recognition is only a more sensitive measure of the underlying trace than recall (a point first made by Anderson & Bower, 1972). By the same token, the dissociations advance the case for theories that assume more than one underlying process in retrieval (generate-recognition theories, for example, or Mandler's [1980] familiarity-plus-retrieval theory). But nowadays recognition and recall are considered as belonging to declarative memory, tested explicitly, and requiring deliberate recollection of the encoding episode. Evidence we used formerly to distinguish them is still valid for just that, but we would not maintain they represent different systems of memory, in the sense of process modularity.

Dunn and Kirsner (1987) have argued that crossed double dissociations can be expected even from single-process models whenever there are two performance measures that are necessarily reciprocal to each other. In the Anderson and Bower (1972) theory of recognition and recall, two separate processes are postulated in order to account for the dissociations mentioned in the last paragraph, pathway tagging and context associations. The dissociation data are indeed consistent with this two-process theory. But they are also consistent with the view that one of these factors is just what is left over after the other is used. Anything that then increases one of the factors would necessarily have to reduce the other. If going from incidental to intentional learning increases pathway tagging, for example, it would have to reduce the importance of what remains, perhaps context associations, without affecting this latter factor directly. Single dissociations, acording to Dunn and Kirsner, are even more easily dismissed. If a common single factor is included in two tasks, its effective range of action might be more suited to one than another. In the limit, if this factor affects one task reliably, it could fail to affect another because within the context of this second task it is at floor or ceiling. For example, the experimental factor of retention interval seems to have different effective ranges of action on primed fragment completion than it does on explicit recall and recognition (Sloman, Hayman, Ohta, Law & Tulving, 1988).

In summary, the empirical dissociations of the past have not been sufficient for postulating multiple memory systems. I have included information about short-term storage, about consolidation theory, and about recall/recognition differences, to make this point. Now what of the dissociations popular today?

THE PROCEDURAL/DECLARATIVE DISTINCTION

We read much evidence, these days, for multiple memory systems based on dissociations between the relation of different tasks and different independent variables (Squire, 1987; Tulving, 1987). The particularly dramatic reports are of dissociations between explicit and implicit memory in amnesics (Cohen & Squire, 1980; Graf & Schachter, 1985; see chapters in Squire & Butters, 1984). The same two systems have also been dissociated in normal subjects as a function of experimental operations (for example, Jacoby, 1983). Nobody doubts that these dissociations have exciting implications for theory; the only question is the legitimacy of concluding that dissociations show the existence of *different memory systems* (Dunn & Kirsner, 1987; Jacoby, 1983; Roediger, 1984). In this section I will argue that such a conclusion is risky at best. I shall not attempt to review the growing evidence on implicit and explicit tests of memory, however, because I think no such review is needed.

Dissociations within implicit and explicit memory. Besides repeating the cautions I have already mentioned about interpreting empirical dissociations, I shall pause here to cite a particularly sobering context for the procedural/declarative (or perhaps implicit/explicit) subsystem distinction: Roediger and Blaxton (1987) have shown that we should not be so quick to declare that a dissociation of one memory measure from another heralds a distinction between systems of memory. They find that striking empirical dissociations occur when several testing procedures for implicit memory are compared in response to the same independent variable. In the end, a dissociation is a particularly well-behaved and replicable *interaction* (Tulving, 1987) in which one set of tasks responds to a manipulation and another does not, or vice versa. We are all used to getting

theoretical mileage from interactions, of course, but additional arguments are needed to defend a distinction between memory systems in the sense of processing modularity.

Dunn and Kirsner (1987) maintain that dissociations, and even double dissociations, are not logically acceptable grounds for distinguishing mental processes. They suggest the method of "reversed association" as a more defensible pattern on which to distinguish process. A reversed association is essentially a nonmonotonic relationship between two tasks across conditions and subjects.

At the least, the preceding review of two-process memory theory should have made clear that proposing distinct processing on the basis of even orderly empirical dissociations is premature. Now we are seeing strong and fascinating empirical dissociations between implicit and explicit tests of memory. As we seek to interpret these, perhaps we should be cautious in suggesting two underlying memory systems, such as procedural and declarative memory.

Two Systems or a modular component? In relation to procedural and declarative memory, we talk as if two systems have been isolated, but really there is only one element—the declarative encoding of temporal context—that is separate from all the other diverse procedural formats. Each of the latter is "stored" in its processing locus in the brain. Procedural memory is really an umbrella term for processing residues of all sorts depending on the mode of original information processing. The structure of Schneider and Detweiler's (1987) recent "connectionist/control" model seems to recognize some of these attitudes: They propose processing modules assorted by principles of vertical modularity, that is, assorted by processing formats—coding modules—such as visual, auditory, phonetic, semantic, lexical, and so on. Among these is a module representing context. In on-line processing, the connections between this context module and other processing centers constitutes attention. Although I disagree with Schneider and Detweiler on the necessity of special assumptions for a system of buffer storage, the survival of contextual connections in the short- and long-term will very plausibly comment on two of our concerns in this essay. First, the uninterrupted pace of contextual change could provide the constancy that gives "primary memory" its Jamesian recollective experience of belonging to the conscious present. Second, the source of classical amnesia may be understood by virtue of a special vulnerability, or fragility in the sense of neo-consolidation theory, of the contextual connections with other aspects of the processing system. The argument that amnesia could not be a disturbance of consciousness is a well-understood proviso in the theory of amnesia —amnesics show no obvious signs of not being aware of the world around them as they go about their lives. It is the survival of traces connecting this awareness with the processing systems used in the past that may be disturbed.

This does not sound to me like a distinction between procedural and declarative memory as two systems, as such. It sounds like one element of normal memory—the knowledge that *that* processing occurred in *that* context—is compromised. On the other hand, if we have a model of memory that is vertically modular, each of quite a few coding formats distinct from the others, then loss of context connections, would be just the loss of information of one among many specialized codes.

REFERENCES

Anderson, J. R., & Bower, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review, 79*, 97-123.

Baddeley, A. D. (1983). Working memory. *Philosophical Transactions of the Royal Society of London, B, 382*, 311-324.

Baddeley, A. D. (1986) *Working memory*. New York: Oxford University Press.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47-89). New York: Academic Press.

Baddeley, A. D., & Hitch. G. J. (1977). Recency revisited. In S. Dornic (Ed.), *Attention and performance 6* (pp. 647-667). Hillsdale, NJ: Lawrence Erlbaum Associates.

Baddeley, A. D., & Scott, D. (1971). Short-term forgetting in the absence of proactive inhibition. *Quarterly Journal of Experimental Psychology, 23*, 275-283.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior, 14*, 575-589.

Bennet, R. W. (1975). Proactive interference in short-term memory: Fundamental forgetting processes. *Journal of Verbal Learning and Verbal Behavior, 14*, 123-144.

Bjork, R. A., & Whitten, W. B. (1974). Recency-sensitive retrieval processes in long-term free recall. *Cognitive Psychology, 6*, 173-189.

Broadbent, D. E. (1958). *Perception and communication*. New York: Pergamon.

Brown, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology, 10*, 12-21.

Case, R., Kurland, M. D., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology, 33*, 386-404.

Chorover, S. L. (1976). An experimental critique of the "consolidation studies" and an alternative "model systems" approach to the biophysiology of memory. In M. R. Rosenzweig & E. L. Bennett (Eds.), *Neural mechanisms of learning and memory* (pp. 561-582). Cambridge Mass. MIT Press.

Chorover, S. L., & Schiller, P. H. (1965). Short-term retrograde amnesia in rats. *Journal of Comparative and Physiological Psychology, 59*, 73-78.

Cofer, C. N, & Davidson, E. H. (1968). Proactive interference in STM for consonant units of two sizes. *Journal of Verbal Learning and Verbal Behavior, 7,* 268-270.

Cohen, N. J., & Squire, L. R. (1980). Preserved learning and retention of pattern-analysing skill in amnesia: Dissociation of knowing how and knowing that. *Science, 210,* 207-210.

Cohen, R. L., & Sandberg, T. (1977). Relations between intelligence and short-term memory. *Cognitive Psychology, 9,* 534-554.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: a framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11,* 671-684.

Crowder, R. G. (1969). Behavioral strategies in immediate memory. *Journal of Verbal Learning and Verbal Behavior, 8,* 524-528.

Crowder, R. G. (1976). *Principles of learning and memory.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Crowder, R. G. (1982). The demise of short-term memory. *Acta Psychologica, 50,* 291-323.

Crowder, R. G., & Greene, R. L. (1987a). The context of remembering. In D. S. Gorfein & R. R. Hoffman (Eds.), *Memory and learning: The Ebbinghaus centennial conference.* (pp. 191-199) Hillsdale NJ: Lawrence Erlbaum Associates.

Crowder, R. G. & Greene, R. L. (1987b). On the remembrance of times past: The irregular technique. *Journal of Experimental Psychology: General , 116,* 265-278.

Crowder, R. G., & Morton, J. (1969). Precategorical acoustic storage (PAS). *Perception and Psychophysics, 5,* 365-373.

Daneman, M., & Carpenter, P. A. (1983). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19,* 450-466.

Dunn, J. C., & Kirsner, K. (1987). *Discovering functionally independent mental processes: The principle of reversed association.* Unpublished manuscript.

Ekstrand, B. R. (1972). To sleep, perchance to dream (about why we forget). In C. P. Duncan, L. Sechrest, & A. W. Melton (Eds.), *Human memory: Festschrift for Benton J. Underwood* (pp. 59-82). New York: Appleton Century Crofts.

Ellis, N. C., & Hennelly, R. A. (1980). A bilingual word-length effect: Implications for intelligence testing and the relative ease of mental calculation in Welsh and English. *British Journal of Psychology, 71,* 43-52.

Empson, J. A. C., & Clarke, P. R. E. (1970). Rapid eye movements and remembering. *Nature, 227,* 287-288.

Estes, W. K., (1972). An associative basis for coding and organization in memory. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory.* Washington, DC: Winston & Sons.

Fisher, R. P., & Craik, F. I. M. (1977). Interaction between encoding and retrieval operations in cued recall. *Journal of Experimental Psychology: Human Learning and Memory, 3,* 701-711.

Fodor, J. A. (1983) *Modularity of mind.* Cambridge, MA: MIT Press.

Fuchs, A. H., & Melton, A. W. (1974). Effects of frequency of presentation and stimulus length on retention in the Brown-Peterson paradigm. *Journal of Experimental Psychology, 103,* 629-637.

Gerard, R. W. (1949). Physiology and psychiatry. *American Journal of Psychiatry, 105,* 161-173.

Glanzer, M. (1972). Storage mechanisms in recall. In G. H. Bower & J. T. Spence (Eds.), *The psychology of learning and motivation* (Volume 5, pp. 129-193). New York: Academic Press.

Glenberg, A. M. (1987). Temporal context and memory. In D. S. Gorfein & R. R. Hoffman (Eds.) *Memory and learning: The Ebbinghaus centennial conference* (pp. 173-190). Hillsdale NJ: Lawrence Erlbaum Associates.

Glenberg, A. M., Bradley, M. M., Stevenson, J. A., Kraus, T. A., Tkachuk, M. J., Gretz, A. L., Fish, J. F., & Turpin, B. A. M. (1980). A two-process account of long-term serial position effects. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 355-369.

Glenberg, A. M., & Swanson, N. C. (1986). A temporal distinctiveness theory of recency and modality effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition. 12,* 3-24.

Gold, P. E., & McGaugh, J. L. (1984). Endogenous processes in memory consolidation. In H. Weingartner & E. S. Parkers (Eds.), *Memory consolidation: Psychobiology of cognition* (pp. 65-84). Hillsdale NJ: Lawrence Erlbaum Associates.

Gorfein, D. S. (1987). STM & discriminability. In D. S. Gorfein & R. R. Hoffman (Eds.) *Memory and learning: The Ebbinghaus centennial conference* (pp. 153-172). Hillsdale, NJ: Lawrence Erlbaum Associates.

Graf, P., & Schacter, D. L. (1985). Implicit and explicit memory for new associations in normal and amnesic subjects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 501-518.

Greene, R. L. (1986). Sources of recency effects in free recall. *Psychological Bulletin, 99,* 221-228.

Hebb, D. O. (1949). *Organization of behavior.* New York: Wiley.

Hilgard, E. R., & Bower, G. H. (1975). *Theories of learning* (4th ed.). Englewood Cliffs, NJ: Prentice-Hall.

Hitch, G. J. (1985) Short-term memory and information processing in humans and animals: Towards an integrative framework. In L. G. Nilsson & T. Archer (Eds.) *Perspectives on learning and memory.* Hillsdale, NJ: Lawrence Erlbaum Associates..

Hockey, G. R. J., Davies, S., & Gray, M. W. (1972). Forgetting as a function of sleep at different times of day. *Quarterly Journal of Experimental Psychology, 24,* 386-393.

Huttenlocher, J., & Burke, D. (1976). Why does memory span increase with age? *Cognitive Psychology, 8,* 1-31.

Jackendoff, R. (1987). The modularity of the computational mind. In R. Jackendoff, *Consciousness and the computational mind* (Chapter 12). Cambridge, Bradford/MIT.

Jacoby, L. L. (1983). Remembering the data: Analysing interactive processes in reading. *Journal of Verbal Learning and Verbal Behavior, 22,* 485-508.

James, W. (1890). *Principles of psychology.* New York: Holt.

Jones, D. M. (1979). Stress and memory. In M. M. Gruneberg & P. E. Morris (Eds.), *Applied problems in memory* (pp. 185-214). London: Academic Press.

Jost, A. (1897). Die Associazionfestigkeit in ihrer Abhangigkeit von der Verteilung der Weiderholungen. *Zeitschrift fur Psychologie, 14,* 436-472.

Keppel, G., & Underwood, B. J. (1962). Proactive inhibition in the short-term retention of single items. *Journal of Verbal Learning and Verbal Behavior, 1,* 153-161.

Kolers, P. (1975). Specificity of operations in sentence recognition. *Cognitive Psychology, 7,* 289-306.

Landauer, T. K. (1974). Consolidation in human memory: Retrograde amnestic effects of confusable items in paired-associate learning. *Journal of Verbal Learning and Verbal Behavior, 12,* 119-131.

Landauer, T. K. (1977). Remarks on the detection and analysis of memory deficits. In E. M. Birnbaum & E. S. Parker (Eds.), *Alcohol and human memory* (pp. 23-42). Hillsdale, NJ: Lawrence Erlbaum Associates.

Lee, C. L., & Estes, W. K. (1981). Item and order information in short-term memory: Evidence for multilevel perturbation process. *Journal of Experimental Psychology: Human Learning and Memory, 7,* 149-169.

Loess, H. (1964). Proactive inhibition in short-term memory. *Journal of Verbal Learning and Verbal Behavior, 3*, 362-368.

Mandler, G. (1980). Recognition: The judgment of previous occurrence. *Psychological Review, 87*, 252-271.

McGaugh, J. L., & Gold, P. E. (1974). Conceptual and neurobiological issues in studies of treatments affecting memory storage. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 233-264). New York: Academic Press.

McGeoch, J. A. (1942). *The psychology of human learning*. New York: Longmans Green & Co.

Miller, G. A. (1956). The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81-97.

Miller, R. R., & Marlin, N. A. (1984). The physiology and semantics of consolidation. In H. Weingartner & E. S. Parkers (Eds.), *Memory consolidation: Psychobiology of cognition* (pp. 85-110). Hillsdale NJ: Lawrence Erlbaum Associates.

Miller, R. R., Springer, A. D. (1973). Amnesia, consolidation, and retrieval. *Psychological Review, 80*, 69-79.

Murdock, B. B., Jr. (1960). The distinctiveness of stimuli. *Psychological Review, 67*, 16-31.

Noyd, D. E. (1965, June). *Proactive and intra stimulus interference in short-term memory for two-, three, and five-word stimuli*. Paper presented at meeting of the Western Psychological Association, Honolulu.

Peterson, L. R., & Peterson, M. J. (1959). Short-term retention of individual items. *Journal of Experimental Psychology 61*, 12-21.

Pollack, I., Johnson, I. B., & Knaff, P. R. (1959). Running memory span. *Journal of Experimental Psychology, 57*, 137-146.

Posner, M. I., & Konick, A. W. (1966). On the role of interference in short-term retention, *Journal of Experimental Psychology, 72*, 221-231.

Reisberg, D., Rappaport, I., & O'Shaughnessy, M. (1984). Limits of working memory: The digit digit-span. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 10*, 203-221.

Ribot, T. (1881). *Les maladies de la memoire*. Paris: Germer Baillere.

Roediger, H. L., III (1984). Does current evidence from dissociation experiments favor the episodic/semantic distinction? *Behavioral and Brain Sciences, 7*, 252-254.

Roediger, H. L., III, & Blaxton, T. A. (1987). Retrieval modes produce dissociations in memory for surface information. In. D. S. Gorfein & R. R. Hoffman (Eds.), *Memory and learning: The Ebbinghaus centennial conference* (pp. 349-379). Hillsdale, NJ: Erlbaum.

Russell, W. R., & Nathan, P. W. (1946). Traumatic amnesia. *Brain, 69*, 280-300.

Salamé, P., & Baddeley, A. D. (1982). Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of Verbal Learning and Verbal Behavior, 21*, 150-164.

Samuel, A. G. (1978). Organization versus retrieval factors in the development of digit span. *Journal of Experimental Child Psychology, 26*, 308-319.

Schneider, W., & Detweiler, M. (1987). A connectionist/control architecture for working memory. In G. H. Bower (Ed.), *The psychology of learning and memory, Volume 21* (pp. 53-119). New York: Academic Press.

Sloman, S. A., Hayman, C. A. G., Ohta, N., Law, J., & Tulving, E. (1988). Forgetting in primed fragment completion. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Squire, L. R. (1987). *Memory and brain*. New York: Oxford University Press.

Squire, L. R., & Butters, N. (Eds.), (1984) *Neuropsychology of memory*. New York: Guilford Press.

Squire, L. R., Slater, P. C., & Chace, P. M. (1975). Retrograde amnesia: Temporal gradient in very long-term memory following electroconvulsive therapy. *Science, 187*, 77-79.

Tulving, E. (1983). *Elements of episodic memory*. New York: Oxford University Press.

Tulving, E. (1987). Introduction: Multiple memory systems and consciousness. *Human Neurobiology, 6*, 67-80.

Turvey, M. T., Brick, P., & Osborn., J. (1970). Proactive interference in short-term memory as a function of prior-item retention interval. *Quarterly Journal of Experimental Psychology, 22*, 142-147.

Tzeng, O. J. L., (1973). Positive recency effect in delayed free recall. *Journal of Verbal Learning and Verbal Behavior, 12*, 436-439.

Watkins, M. J. (1977). The intricacy of memory span. *Memory & Cognition, 5*, 529-534.

Watkins, M. J., & Peynircioglu, Z. F. (1983). Three recency effects at the same time. *Journal of Verbal Learning and Verbal Behavior, 22*, 375-384.

Waugh, N. C., & Norman, D. A. (1965), Primary memory. *Psychological Review, 72*, 89-104.

Weingartner, H. (1984). Psychobiological determinants of memory failures. In L. R. Squire and N. Butters (Eds.), *Neuropsychology of memory* (pp. 205-212). New York: Guilford Press.

Whitten, W. B., & Bjork, R. A. (1972). *Test events as learning trials: The importance of being imperfect*. Paper presented at the Midwestern Mathematical Psychological meeting, Bloomington, IN.

Wickelgren, W. A. (1977). *Learning and memory*. Englewood Cliffs, NJ: Prentice-Hall.

Wickelgren, W. A. (1979). Chunking and consolidation: A theoretical synthesis of semantic networks, configuring in conditioning, S-R versus cognitive learning, normal forgetting, the amnesic syndrome, and the hippocampal arousal system. *Psychological Review, 86*, 44-60.

Woodworth, R. S. (1938). *Experimental psychology*. New York: Holt.

Wright, J. H. (1967). Effects of formal interitem similarity and length of retention interval on proactive inhibition in short-term memory. *Journal of Experimental Psychology, 75*, 366-395.

Zechmeister, E. B., & Nyberg, S. E. (1982). *Human memory: An introduction to research and theory*. Monterey, CA: Brooks/Cole.

FOOTNOTES

[*] In H. L. Roediger, III & F. I. M. Craik, (Eds.), *Varieties of memory and consciousness* (pp. 274-294). Hillsdale, NJ: Lawrence Erlbaum Associates (1989).

[†] Yale University, Department of Psychology.

[1] Crowder and Greene (1987a, b) have commented on a possible error in conceptualizing the *modality effect* in this way, but my concern here is with the general idea that recency results from temporal discriminability of items at the end of a series.

[2] I say "largely" with cause: Salamé and Baddeley (1982) have subdivided the articulatory loop into storage and processing components, which are examples of processing modularity within a form of coding.

[3] Others (Hilgard & Bower, 1975, pp. 508-516) have reviewed essentially the same literature and reached the conclusion that consolidation is the concept.

Haskins Laboratories Status Report on Speech Research
1990, SR-105/106, 83-92

Representation and Reality: Physical Systems and Phonological Structure*

Catherine P. Browman and Louis Goldstein†

Pierrehumbert's contribution lays out, in lucid fashion, the problems inherent in relating different kinds of representations of speech, in particular, representations that have as their goal elucidating systems of contrast and combination of units (the phonological), and representations that have as their goal precise physical descriptions (the phonetic). She argues that, in resolving these problems, first, phonetics and phonology cannot be treated in isolation from each other; second, properties of the real world must be seen as constraining the relation between phonetics and phonology; and third, a one-to-one correspondence between quantitative and phonetic, on the one hand, and abstract and phonological, on the other, cannot be maintained—mental representations can be quantitative, and physical representations abstract. On all these points, we agree with Pierrehumbert.

Pierrehumbert also maintains that phonology and phonetics involve disparate representations, drawing upon several dichotomies to help characterize the differing kinds of terrain for which these different representations are assumed to be suitable guides. One of these dichotomies is *cognitive* vs. *physical*. Another is *qualitative* vs. *gradient*. The phonetic domain is seen as physical and gradient and is described by "continuous" mathemetics (the calculus). The phonological is seen as cognitive and (for the most part, but not completely) qualitative and is described by formal languages, rather than calculus.

Our thanks to Carol Fowler, Peter Ladefoged, Janet Pierrehumbert, Elliot Saltzman, Michael Studdert-Kennedy, and Michael Turvey for critiquing an earlier version of this paper. Any remaining errors are of course our responsibility alone. This work was supported by NSF grant BNS 8820099 and NIH grants HD-01994 and NS-13617 to Haskins Laboratories.

While these dichotomies are superficially plausible, recent work on self-organization in complex biological and physical systems (e.g., Haken, 1977) can be taken as a lesson that the rich structuring shown by these systems can be understood when the importance imparted to such dichotomies is abandoned. In the paragraphs below, we first give some examples (from speech and elsewhere) that incline us against enforcing these kinds of dichotomies. We then introduce an alternative suggestion for speech representations that can serve the different goals that Pierrehumbert identifies. In this alternative, the goals are served by macroscopic and microscopic properties of one and the same system. We conclude by exemplifying this approach using speech error data.

1. Cognitive and physical need not be distinct

Pierrehumbert argues that phonology is, *prima facie*, cognitive, because phonological contrast is the basis for the association of form and meaning, which in turn must be part of an individual's cognitive structure. However, because she also assumes that the cognitive component is qualitative and discrete, and further that the physical cannot be qualitative (except as "analyzed" by a cognitive system), the cognitive and the physical must be different (and representationally incommensurate). However, as we will show below, the physical world is both gradient and qualitative, and there is no reason why the cognition cannot be attuned to those qualitative aspects of the real world (in this case the activity of talkers and listeners), rather than imposing a discrete, qualitative order on an otherwise homogeneously continuous world. Indeed, the entire research program of "direct realism" (e.g., Fowler, 1980; Fowler, Rubin, Remez, & Turvey, 1980; Gibson, 1966, 1979; Turvey, 1977) proceeds

from the assumption that human beings' perceptions of the world are, in fact, true of the real world, and investigates how the perceptual information specifies properties of the real world.

Recent research on coordinated movement provides a number of challenges to the view that cognitive and physical systems are distinct and incommensurate. This research (e.g., Kelso & Tuller, 1984; Kugler & Turvey, 1987; Turvey, 1990) shows that in order to account for coordinated movement, the physical properties of an organism must play a central role, and that coordination is accomplished through an interaction of cognitive and neural activity with the principles of physical self-organization. A particularly striking demonstration of this can be seen in the nature of phase transitions between alternate modes of coordinating rhythmic movements (Kelso, 1984; Kelso & Scholz, 1985). When two index fingers are oscillated at the same frequency, two phasing modes are possible, symmetric and anti-symmetric. If a person starts oscillating in the antisymmetric mode and is asked to speed up, eventually there is an abrupt (and involuntary) shift to the symmetric mode. (The reverse is not true). This phase shift has been modeled by Haken, Kelso, and Bunz (1985) using a potential function that predicts the stability of particular phase relations as a function of the oscillation frequency of the components. A number of empirical details of the phenomenon (such as the appearance of fluctuations in phase as the critical point is reached) support this physical stability analysis and the hypothesis that the sudden change in phase is a physical bifurcation related to increased instability of the antisymmetric mode.

Now in this simple kind of task, the fact that the results can be predicted on the basis of a physical systems analysis may not seem particularly surprising. After all, the human body *is* a physical system (whatever else it is), and it is not hard to convince oneself that the two fingers are mechanically (or at least neurally) coupled to one another. A common intuition here is that it is just somehow "more difficult" to move the fingers in the antisymmetric mode, and at high frequencies, the subject simply gives up. Given such an analysis, this task might not have much to say about the role of cognition. Recently, however, Schmidt, Carello, and Turvey (1990) have extended this method to the case where the two limbs being coordinated belong to two different people. The subjects face each other and are instructed to swing their legs in synchrony (either in or out of phase). Exactly the same results were obtained as in the within-person case—an abrupt shift to the symmetric mode as frequency was increased, with all the hallmarks of a physical bifurcation. So what we have is a system that is behaving in a way that is well explained by physical systems, yet there is no mechanical (or hard-wired neural) coupling between the oscillating elements in this case. The coupling must be, as the authors conclude, *informational*. To the extent that we think of this kind of information as "cognitive" (which in a broad sense it must be; certainly "visual information processing" has traditionally been so considered), cognitive activity actually functions to couple the elements of a (single) physical system. It is hard to imagine a tighter interpenetration of the cognitive and the physical.

Of course, one may want to object that the sense of cognitive here is not the same one that Pierrehumbert is employing. Specifically, she is referring to a kind of introspectively available cognition that is not specifically perceptual or motoric. But there is still another moral lurking here. The coupled swinging legs constitute a physical system, but it is one that is softly assembled. That is, the two people can sit and swing their legs and watch each other without *intending* to synchronize their swings. Without the intentionality, the complex system is never actually assembled from the pieces (although there might be some spontaneous tendency to synchronize anyway). Thus, the (clearly cognitive) act of intending to synchronize provides the boundary conditions under which the self-organization of the physical system takes place (Kugler & Turvey, 1987; Turvey, 1990).

2. Physical systems are simultaneously gradient and qualitative

Perhaps the most dramatic examples of the simultaneously gradient and qualitative nature of the physical world are the instances of "self-organization" that have attracted attention in recent years (e.g., Madore & Freedman, 1987; Prigogine & Stengers, 1984). These demonstrations show that, under the right conditions, complex, qualitatively distinct forms may spontaneously emerge in previously homogeneous, undifferentiated media. For example, with the right concentration of reagents, a quiescent petrie dish will suddenly display colored concentric circles and rotating spirals, which eventually dominate the entire surface and then die out (the Belousov-Zhabotinskii reaction). One can describe what is going on in the petrie dish in different ways. From one point of

view, it is clearly gradient—there is spatial and temporal variation in the concentration of particular ions that can be described in terms of the relevant (continuous) differential equations. From another point of view there are a number of distinct observable qualitative forms (e.g., circles and spirals). These are descriptions of the same physical system, differing only in the "grain" employed.

Self-organized systems, like the chemical oscillators described above, provide another window onto the relation between gradient and qualitative properties of physical systems. The reaction described above (and others, see Prigogine & Stengers, 1984) occurs only when the concentration of some of the reagents is within some critical limits. Thus, while concentration is a gradient quantity, the concentration continuum is inherently partitioned into regions which show qualitatively distinct behavior (quiescence vs. ring-formation). This partitioning into distinct long-term behaviors, or *bifurcation* as it is sometimes called, is a common property of physical systems. Even an equation as simple as (1), used to describe population dynamics (May & Oster, 1976, in Gleick (1987); Hofstadter, 1981), shows such properties.

$$x_{next} = rx(1 - x) \quad (1)$$

(1) is iterated to get the population of each successive generation. Depending on the value chosen for r, the long-term behavior of the system will differ. At low values, the system settles down to a single value for x in successive iterations. As r increases, the behavior abruptly changes qualitatively, yielding a stable alternation of two values for x. This period-doubling bifurcation will occur again at some point as the value of r is increased further, yielding a cycling among four values. The doubling continues until at some point the system becomes chaotic, and does not show any predictable pattern of repetition. Thus, while equation (1) is defined in the world of continuous mathematics, it provides a landscape of discretely different behaviors as a function of the (gradient) value of its parameter. The potential for such discrete, qualitative behavior is always lurking in systems that include non-linear terms (cf. Thompson & Stewart, 1986), even quite simple ones, such as (1). Only in strictly linear systems is the gradience of the parameter space mirrored by a relatively undifferentiated landscape of system behavior (and even then, not always). Complex systems in the real world, however, involve non-linearity and qualitatively distinct behaviors.

There are various ways in which speech, as a physical system, can be seen as simultaneously gradient and qualitative. One familiar example is Stevens' quantal theory (Stevens, 1972, 1989). This theory holds that while it is possible to describe constrictions within the vocal tract in terms of continuous (geometric) parameters, the acoustic (and auditory) properties of the sound produced by the vocal tract are such that the continuous parameter space is effectively partitioned into discrete regions. Within each region, the auditory properties are stable (they don't vary greatly as a function of small changes in the articulation), and qualitatively distinct from auditory properties associated with other regions. These regions are thus seen as the basis for contrast (distinctive features, for Stevens). To the extent to which this theory is correct, then speech, like some of the other examples of complex systems described above, is a physical system that is intrinsically both gradient and qualitative.

A second example involves the characterization of articulatory trajectories during speech. As Pierrehumbert notes, the articulators are constantly in motion, and thus the trajectories can be described in a gradient fashion. However, as a number of researchers have hypothesized and found, it is possible to model the time-varying motion of the articulators (during a particular speech gesture) using an invariant dynamical specification (Browman & Goldstein, 1985, 1990; Beckman, Edwards, & Fletcher, in press; Fowler et al., 1980; Kelso, Vatikiotis-Bateson, Saltzman & Kay, 1985; Ostry and Munhall, 1985; Vatikiotis-Bateson, 1988). That is, even though the articulators are moving, the underlying dynamical system that gives rise to this motion is not varying over time. There is a discrete interval of time during which this (temporally invariant) regime for a particular gesture is active. Thus, once again, the speech system exhibits behavior that is at one and the same time gradient and qualitative.

3. Macroscopic and microscopic properties of phonological structure

While the cognitive-physical and qualitative-gradient dichotomies do not seem to us to be useful, the differing representational goals that Pierrehumbert identifies as phonological vs. phonetic must be satisfied in some way. We would like to suggest that these can be construed as macroscopic and microscopic properties of a single complex system. Thus, properties such as contrast and paradigmatic and syntagmatic relations are coarse-grained, macroscopic, relational properties

that hold among the system's units. The precise articulatory and acoustic characterizations of these units (and their variation) are fine-grained, microscopic properties.

This macro-micro perspective does more than just provide a different name for a familiar distinction. The fact that macroscopic and microscopic properties simultaneously characterize the same complex (physical) system has substantive consequences. Recent work has begun to explore the general properties of cooperativity among system components, and of the linkage between patterns at different "scales" (Kugler & Turvey, 1987; Schoner & Kelso, 1988). An important characteristic of complex physical systems showing "self-organization" is that there is an interaction (or reciprocity) between the microscopic and macroscopic properties of the system (Kugler & Turvey, 1987). The nature of the microscopic units, or atomisms, affects the possible stable macroscopic structures, and the macroscopic organization affects the behavior of microscopic units.

Kugler and Turvey (1987) present examples of such micro-macro cooperativities. For example, they show how nest-building in insects can be analyzed in this way. Macroscopic chemical gradients *originate in* the behavior of individual insects (whose deposits contain pheromones), but these gradients also *constrain* the activities of the insects (causing them to deposit in high-pheromone density locations which results in the formation of macroscopic pillars and arches). This example was also used by Lindblom, MacNeilage, and Studdert-Kennedy (1983), in order to show how certain phonological units could be derived from properties of speech. However, it is important to see that the micro-macro constraints are typically *reciprocal*.

For example, Kugler and Turvey present experiments that demonstrate micro-macro reciprocity in human coordinated movement. When individuals are asked to swing a pendulum from the wrist, the preferred "comfort mode" frequency depends on the size of the pendulum. When asked to swing two differently-sized pendula (one in each wrist) in absolute coordination, the observed frequency does not correspond to either single frequency, but it turns out to correspond to the frequency of a macroscopic virtual system, that treats that the two pendula as rigidly coupled. The macroscopic properties of the system are determined by the microscopic properties of the components (their natural frequencies), but the coupled system constrains, in turn, the individual wrist-pendulum systems, pushing them away from their own natural frequencies (see also Turvey, Rosenblum, Schmidt, & Kugler, 1986). Schmidt (1988, discussed in Turvey, Saltzman, & Schmidt, 1991) shows that such effects hold even when it is two different individuals swinging the two pendula, so that the coupling is informational.

Applying this macro-micro perspective to phonological structure would involve showing that the contrastive and combinatoric properties (the ones Pierrehumbert calls phonological) arise out of the microscopic (articulatory and acoustic) properties of the individual atomisms (such as articulatory gestures), and, in turn, that the macroscopic properties constrain the details of the units. This perspective predicts that reciprocity between the grains of description should be the rule, not the exception. Such reciprocity, or mutual constraint, would come as a surprise to the phonological and phonetic "imperialists" that Pierrehumbert attacks, but fits more comfortably with her outlook that a theory encompassing both domains, as well as their relations, is necessary to a full understanding of phonology and phonetics. Likewise, the syntactic view of phonetic/phonology relations, that Pierrehumbert rejects, runs afoul of such reciprocities. However, it seems to us that the semantic type of mapping which she proposes is also not a good analogy here. As Pierrehumbert notes, one mapping called "semantic," the lexical sound-meaning correspondence, is too arbitrary to capture what is going on in the phonology-phonetics relation. Even if the semantic mapping in question is that between a concept and its real-world extensions (such as the concept DOG and the set of dogs in the real world), it differs substantively in its possibilities for macro-micro reciprocities from the phonology-phonetics relation. While both concepts and phonological structure may be dependent on real-world conditions, their potential for affecting real world properties differs considerably. That is, although one's concept of DOG may affect one's relation with a real-world dog (e.g., patting or running from it), the concept does not have the same potential to constrain the nature of that dog that phonological structure does to constrain the properties of speech.

While macro-micro reciprocity has not been demonstrated conclusively for language at this point, it is possible to find a variety of examples of such reciprocity. One such example can be seen in the organization of English vowels into the paradigmatic system revealed in a series of "chain shifts" identified by Labov, Yaeger, and Steiner (1972). In these sound changes in progress, subsets of vowels show coordinated patterns of

movement along particular "tracks" in the vowel space. In particular, tense (and ingliding) vowels tend to raise, so that low vowels become mid, and mid vowels become high. These constrained (and apparently universal across dialect) patterns of movement and the paradigmatic relations that they reveal clearly constitute a macroscopic property of the English vowel system (and one that would be captured as part of the phonology, in most accounts). It is possible, however, to see reciprocity between this structure and the microscopic properties of the vowel units themselves.

First, Goldstein (1983) has shown that it is possible to derive the tracks along which vowels move during shifts on the basis of articulatory-acoustic relations. Random articulatory perturbations of model vowels results in acoustic variation that is directed along the dimensions that are involved in chain shifts (principally, the high-low dimension). Thus, the microscopic properties of these vowels gives rise to the layout of the macroscopic tracks.

Second, there is also evidence that the macroscopic state of the vowel system also constrains the microscopic properties of individual vowels. Labov (1972) showed that while part of an "active" (macroscopic) shift (which might span 50-75 years), a vowel shows large, regular (microscopic) context effects in speakers' productions. Before and after the shift, however, the context effects are much smaller. Thus, the macroscopic dynamical state of the system (actively moving vs. static) constrains the microscopic properties of the vowels. This difference in amount of contextual variation at different macroscopic stages provides another kind of support for the complex self-organized system analysis. When such systems are pushed away from particular stable configurations, they typically show an increase in fluctuations, that is, variations in behavior that differ from the stereotypic pattern (Turvey et al., 1986). As a critical point for a nonequilibrium phase transition is reached (as in the finger oscillation example discussed earlier), such fluctuations become extremely large, and provide part of the signature of the change in state (Schoner & Kelso, 1988). The context variation of individual vowel productions can be viewed as just such fluctuations. When the vowel system is at a critical "point," (where a point at this time scale can last many years), increased fluctuations (variation) can be observed.

A different kind of macro-micro linkage may be found in the detailed language-particular differences in the physical properties of the same phonological or categorial phonetic structures (Keating, 1985, 1990). In at least some cases, these microscopic language differences are correlated with the macroscopic structure of contrasts. Let us take one of the examples that Pierrehumbert gives, the fact that the precise frontness of high front unrounded vowels may vary from language to language. Wood (1982) has shown that a prepalatal constriction location is preferred for /i/ in languages which contrast /i/ and /y/, whereas a midpalatal location may be found in languages without the rounding contrast. He then presents modeling results that suggest that this difference is functional—a prepalatal location yields a greater acoustic differentiation of /i/ and /y/. Thus it appears that the macroscopic property of contrast is constraining microscopic properties of the units. Ladefoged (1982) presents a number of examples of this kind, in which language differences in details of production can be related to presence or absence of certain contrasts.

Finally, it has been shown that the amount of contextual variation evidenced by a given phonetic unit may vary. This allowable "region" for a given unit has been modeled by Keating (1988) as a "window," and by Manuel (1990) and Manuel and Krakow (1984) as a "target area." These latter two papers have related the size of the target areas for vowels to the number of contrastive vowels in the system, showing that the areas are smaller when there are more vowels in the system. Again, this seems to suggest an interaction between macroscopic and microscopic system properties.

In the last two examples, the effect of contrast on phonetic units does not provide evidence of reciprocity, *per se*. The effects have seemed only to propagate from the macroscopic to the microscopic. However, there have been a number of attempts to show how the qualitative properties of contrast and combination of phonological systems arise from, or at least are constrained by, the articulatory and acoustic properties of speaking (Stevens' quantal theory—Stevens, 1972, 1989; Lindblom's theory of adaptive dispersion theory—Lindblom & Engstrand, 1989; Lindblom, MacNeilage, & Studdert-Kennedy, 1983; Ohala's vocal tract constraints—Ohala, 1983).While none of these attempts is completely successful, it seems clear that the formation of phonological systems is at least partly molded by the articulatory and acoustic properties of talking. Thus, this is another instance of macro-micro reciprocity: systems of contrast are founded on the microscopic properties of talking, but they also constrain microscopic properties.

In the above sections, we have illustrated that treating phonological and phonetic representations as incommensurate, on the basis of dichotomies such as cognitive-physical and qualitiative-gradient, is probably misguided. These distinctions are irrelevant to, or get in the way of, an understanding of complex physical systems. Within the view that phonological structure is a complex, self-organized system, it is possible to acknowledge that different descriptive tools (e.g., symbols vs. equations) are appropriate for different classes of phonological phenomena, while treating the phenomena as differing grains of a single integrated system. In practical terms, when a linguist is describing some regularities for which a single descriptive tool is appropriate, it may be possible (in some circumstances) to ignore other grains of description. However, evidence presented for reciprocity between macroscopic and microscopic properties strongly suggests that when pursuing a complete understanding of phonological structure and the cognitive/physical activity of talkers and listeners, one ignores the complete system at one's peril.

4. Phonological structure, its representation and transcription: The status of the segment

As we have suggested, paradigmatic contrast and syntagmatic combination can be usefully viewed as important macroscopic properties of phonological system structure. As such, they ought be captured in a *representation* of this system structure. While there are many ways in which this could be done, the use of the *segment*, or a *phonemic transcription*, has been very widely employed as a particular hypothesis, identifying the unit of contrast with the unit of combination. However, we would argue that the basis for such units seems to be in their utility as a practical tool rather than in their correspondence to important informational units of the phonological system.

The primarily practical utility of segmental transcriptions is noted by Pierrehumbert (1990) who sees fine phonetic transcription as "a convenience for the researcher attempting a rough organization of his observations," but finds "no evidence that the elements of fine transcription can be viewed as elements of a discrete representation in the mind." Although Pierrehumbert very clearly distinguishes fine transcription from transcription using phonologically distinctive elements, the pragmatic view towards transcription would appear to be extended to all transcription by Ladefoged (1990). He quotes from Abercrombie's (1964) definition of a phonemic transcription as one in which "the smallest possible number of different letters [symbols] ... distinguish unambiguously all words of different sound in the language." Given that an alphabetical transcription system is being used, this means the symbols used are "segments."

It is important not to confuse the units of such a practical descriptive tool, however useful, with qualitative, informational units that function in the (cognitive/physical) phonological system, when viewed from a macroscopic perspective. The criteria for a useful practical representation (such as economy) are different from those relevant to representing theoretically significant aspects of the system. It seems to us, however, that the success of segmental symbol strings as practical devices has inclined many to make just this confusion, and to assume an isomorphism between the units of transcription and the units of the system itself. Thus, just as Pierrehumbert suggests that fine phonetic transcription has no real theoretical status in phonetics, we suggest that there is no reason to assume that representations employing segmental transcriptions have any theoretical status in phonology.

From this perspective, then, phonemes become one particular, linear, local, and symbol-oriented—"segmental"—solution to the necessity of capturing two related kinds of macroscopic information: distinctive aspects of lexical items, and groupings of allophones (whether alternate pronunciations of the same word, or regular variations, that is, restrictions of distributions). It is, however, not a necessary solution; these same facts can be captured with other units, including gestures and constraints on gestures. Indeed, for almost half a century, the unit of distinctiveness has usually been considered to be, not the segment, but the feature. Moreover, recent phonological proposals such as feature geometry (Clements, 1985; McCarthy, 1988) have explicitly separated the paradigmatic and syntagmatic properties that have been traditionally conflated in the segment (when viewed as a feature bundle). In particular, root nodes are only syntagmatic units in these proposals, constraining how features combine, but it is the features that convey contrast, and they can align in various ways with respect to the syntagmatic frames.

From this perspective, the segmental hypothesis can be viewed as being primarily a specific (local and linear) hypothesis about featurally cohesive syntagmatic units. Researchers have also attempted to extend the segment to indicate a linear

chunking of speech, or to subdivide some larger unit such as the syllable. To at least a first approximation, the segment has been useful in dealing with the acoustic signal. That is, localized linear segmentations of the acoustic signal have real validity, at a coarse level of description. However, it appears that the value of the segment even in characterizing the acoustic signal is limited to the kind of rough organization of observations that Pierrehumbert mentions. The segmental approach runs into trouble, for example, in the syntagmatic world of actual utterances, where it is difficult to find acoustic invariance for any single segment, and where the information associated with a segment might in fact not be temporally localized in the region of that segment, but extend throughout the syllable, or even into other syllables.

Nearey (1990) as well as others has attempted to handle this latter problem by redefining the segment as sensitive to information present in an entire VC (or theoretically, an entire VCV). This of course completely destroys the simple physical definition of a segment as a local linear chunking of the acoustic signal. Nearey explicitly disavowed any featural assumptions in this paper, but investigated the hypothesis of the segment as a subsyllabic unit by comparing segmental and "transsegmental" (i.e., diphone VC or CV) models. While the title of the paper might lead the casual reader to think the paper presents evidence that the segment is a unit of speech perception, in fact Nearey found that a V × C bias component was essential, so that a "pure" segmental model was inadequate. Moreover, Nearey did not compare the segmental hypothesis to other hypotheses of subsyllabic units, such as syllable components (onset, nucleus, coda) or gestures. Thus, his analyses assume acoustic information is distributed transsegmentally and support a transsegmental cognitive component, as well as possibly providing evidence for some kind of subsyllabic cognitive unit. (In fact, it appears to us that his results would be consistent with a gestural analysis.)

Nearey (1990) also cited speech production error data as evidence in favor of the relevance of the segment to speakers' behavior. However, we argue that speech production errors provide no evidence for the behavioral relevance of the segment. (Similar conclusions were reached by Roberts, 1975, and Boomer & Laver, 1968).

It has been repeatedly observed that most speech production errors are single feature errors, and therefore an analysis in terms of features (or feature-like entities) is necessary to describe one important aspect of the entire corpus. Beyond this, however, production errors appear to be divided into two categories: non-interaction and interaction errors (Shattuck-Huffnagel, 1986). These categories differ in two ways: whether a featural description is sufficient, and whether the errors are concentrated in word onsets.

Thirty to forty percent of the errors in the MIT corpus fall into the category of non-interaction errors (30%: Shattuck-Huffnagel & Klatt, 1979; 40% of 1984 count: Shattuck-Huffnagel, 1987). For errors in this category, there is no obvious source for the error in the environment (e.g., "the *D*utch publishers" → "the *G*utch publishers": Shattuck-Huffnagel & Klatt, 1979), and therefore a purely featural analysis is sufficient to describe this subset (assuming it maintains the same featural distribution as the corpus as a whole). The errors tend to occur throughout the word, rather than being concentrated in word-onset position (Shattuck-Huffnagel, 1987).

Sixty to seventy percent of the MIT error corpus consists of interaction errors: anticipatory, perseveratory, and exchange errors (e.g., "they cut their *h*air *sh*ort" → "they cut their *sh*air *h*ort": Shattuck-Huffnagel, 1987). In these errors, the two consonants presumed to be causally interacting are much more similar than would be expected from an interaction of purely independent features, and therefore a purely featural analysis is not sufficient to account for this subset of errors. Rather, some kind of featurally cohesive unit is necessary to describe the interaction errors, unlike the non-interaction errors. Although Shattuck-Huffnagel and Klatt (1979) suggested this featurally cohesive unit was the segment, they had not at that time considered other possibilities, such as syllable or word onsets, which have been shown to account for more data in studies that have compared segmental and onset hypotheses (e.g., Shattuck-Huffnagel, 1983; Vitz & Winkler, 1973).

The interaction errors tend to occur in word onsets. In fact, 82% (and as much as 91% of the exchange error subset) occur in word onsets (Shattuck-Huffnagel, 1987). Shattuck-Huffnagel (1987) argued that when a word-initial consonant participates in an interaction error, "it usually does so by virtue of the fact that it is a word onset (p. 37)," even when it is a single consonant. The featurally cohesive unit for interaction errors, then, appears to be the word onset, not the segment. For example, out of 40 word-onset consonant clusters participating in exchange errors, 36 involved the cluster as a whole (e.g., "*br*eathing and *sm*oking" → "*sm*eething and *br*oking": Shattuck-Huffnagel, 1987).

The validity of the analysis of the structural importance of the lexical item in interaction errors is supported by the striking similarities between production interaction errors and lexical retrieval errors. Brown and McNeill (1966) showed that word onsets (and endings) differ from the rest of the word in being recalled more often in the tip-of-the-tongue state. Browman (1978) showed more specifically that, in addition to a general tendency for "gregariousness" (or "stickiness," in Nearey's (1990) terms), word-initial onsets, word-final VCs, and pre-stressed onsets are prominent in lexical retrieval errors. The two onset categories are thus prominent in both lexical retrieval errors (Browman, 1978) and production interaction errors (or at least exchange errors: Shattuck-Huffnagel, 1987). In addition, production interaction errors and lexical retrieval errors are similar in apparently having separable item and order (or filler and slot) components. Finally, Browman (1978) has argued that the prominence structure in lexical retrieval errors is an attribute of the retrieval process rather than of the lexical entry, which (in conjunction with the other similarities) suggests a possible identity between this process and Shattuck-Huffnagel's (1987) first stage process during which the interaction errors are posited to occur. Thus, it seems likely that the patterns observed for the production interaction errors are attributable to the same process that is observed in lexical retrieval errors (except for the difference in word-final prominence), lending support to the analysis of these errors in terms of lexical units.

To recapitulate: Word onsets plus independent features (or feature-like entities) are necessary and sufficient to account for most interaction errors, while features (or feature-like entities) are sufficient to account for the non-interaction errors. That is, cohesion in speech production errors appears to be defined with respect to the word. The featurally cohesive units are not the same everywhere in the word, nor are they segments, or even onsets of syllables. Rather, the cohesive units are the onsets of words. Arguments based on cohesion do not support the segment in production errors.

However, upon occasion errors involving word onsets break the onset into components, both in production interaction errors and in lexical retrieval errors. While such divisions might appear to be evidence for segments, we suggest that instead they are evidence for articulatory gestures. Like segments, gestures have the potential to be independent movable entities, and can combine into higher level units such as onsets, words, etc. Although a complete analysis in terms of gestures would need to be performed for a more nearly definitive statement, it is nevertheless suggestive that of the 40 errors listed in the Appendix of Fromkin (1973) under "Division of Consonant Clusters," approximately 35 can be analyzed as the movement of a single gesture. Moreover, at least two of the archisegments argued for by Stemberger (1983) as psychologically real units can be equated with gestures.

Many of the similarities between targets and errors captured in earlier analyses as featural similarities can also be captured using gestures. Using the pseudo-gestural analysis in Figure 1 of the confusion matrix from Table 2 in Shattuck-Huffnagel and Klatt (1979), the distribution of numbers of gestures differing between the target and error is suggestive of gestural independence (81% 1 gesture, 18% 2 gestures, 1% 3 gestures). And at least 48 (and possibly up to 54) out of 54 "single feature" errors listed in the Appendix of Fromkin (1973) are also single gesture errors (e.g., "*ped*estrian" → "*teb*estrian," exchange of oral gestures). Note that if gestures are indeed a basic unit, then higher level phonological units such as words, onsets, etc., are associations of gestures (gestural "constellations"), although not necessarily in segmentally-sized units as assumed by Nearey (1990).

	p	b	t	d	k	g	f	v	s	z	th	dh	sh	zh	ch	jh	w	y	r	l	h	m	n	ng
"velic"																						nasal		
												oral												
"oral"	stop						fric								affric		approxrho lat					stop		
	lab		alv		vel		lab		alv		inter		pal		pal		lab pal alv alv					lab alv vel		
"glottal"	vl	vd	vl	vd	vl	vd	vl	vd	vl	vd	vl	vd	vl	vd	vl	vd	vd	vd	vd	vd	vl	vd	vd	vd

Figure 1. **Pseudo-gestural analysis of segments in Table 2 of Shattuck-Huffnagel and Klatt (1979).**

The hypothesis that gestural primitives are critically involved in speech errors makes an additional claim. As defined in articulatory phonology (Browman & Goldstein, 1986, 1989), gestural units are simultaneously discrete units of contrast and quantitatively specified units of articulatory action. The analyses in the previous paragraphs made use of the discrete properties of gestures, since most work on speech errors has assumed that they involve misorderings of discrete, qualitative units in the plan for an utterance. In fact, it has been argued that such misorderings occur before the units receive any articulatory instantiation (Shattuck-Huffnagel, 1987). Gestures, however, "always" have quantitative (gradient) articulatory properties, and if they are the units implicated in speech errors, then it should be possible to find evidence of these properties.

Such evidence is provided in a recent study by Mowrey and MacKay (1990). In this study, they recorded muscle activity during experimentally elicited speech errors, where the errors were induced using tongue twisters such as "Bob flew by Bligh Bay." The electrode placement allowed them to examine activity for [l]. For one recording session of this particular tongue twister, 48 of 150 tokens were produced with anomalous tongue muscle activity (outside the normal range of variation). These anomalies involved the insertion of [l] activity at time points where it was not appropriate (e.g. in "Bob" or "Bay") and the diminution of [l] activity in positions where it was expected. Crucially, these errors were graded. The inserted [l] activity showed a continuum from small amounts of activity to a level consistent with an intended [l]. Likewise, diminution of [l] showed a continuum of reduced activity. Overall, only five tokens showed an "all-or-none" change. Such graded activity is consistent with the quantitative characterization of gestures (in fact, reduction of magnitude has been proposed as a general property of gestures in casual speech, Browman & Goldstein, 1987), but is not consistent with a purely discrete pre-articulatory view of these errors.

Yet in another sense, the errors did show qualitative or discrete behavior. The inserted [l] activity was not smeared throughout the sentences, but was localized at very specific points with temporal profiles comparable to those of an intended [l], even though reduced in magnitude. This is also consistent with the discrete nature of gestures. A gesture is a dynamical system, with an invariant parameter set, that is active for a finite interval of time. As far as it is possible to tell from the muscle activity, the anomalous activity involved an inserted discrete unit of this kind, but with variably reduced magnitude. Thus, it appears that dual nature of gestures—discrete and quantitative—may well be crucial in accounting for speech errors.

REFERENCES

Abercrombie, D. (1964). *English phonetic texts*. London: Faber & Faber.

Beckman, M. E., Edwards, J., & Fletcher, J. (In press). Prosodic structure and tempo in a sonority model of articulatory dynamics. In G. Docherty & D. R. Ladd (Eds.), *Papers in Laboratory Phonology II*. Cambridge: Cambridge University Press.

Boomer, D. S., & Laver, J. D. M. (1968). Slips of the tongue, *British Journal of Disorders of Communication*, 3, 1-12.

Browman, C. P. (1978). Tip of the tongue and slips of the ear: Implications for language processing. *UCLA WPP*, 42.

Browman, C. P., & Goldstein, L. (1985). Dynamic modeling of phonetic structure. In V.A. Fromkin (Ed.), *Phonetic linguistics* (pp. 35-53). New York: Academic Press.

Browman, C. P., & Goldstein, L. (1986). Towards an articulatory phonology. *Phonology Yearbook*, 3, 219-252.

Browman, C. P., & Goldstein, L. (1987). Tiers in articulatory phonology with some implications for casual speech. *Haskins Laboratories Status Report on Speech Research*, 92, 1-30.

Browman, C. P., & Goldstein, L. (1989) Articulatory gestures as phonological units. *Phonology*, 6, 201-251.

Browman, C. P., & Goldstein, L. (1990). Gestural specification using dynamically-defined articulatory structures. *Journal of Phonetics*, 18, 299-320.

Brown, R., & McNeil, D. (1966). The 'Tip of the Tongue' phenomenon. *Journal of Verbal Learning and Verbal Behavior*, 5, 325-337.

Clements, G. N. (1985). The geometry of phonological features. *Phonology Yearbook*, 2, 225-252.

Fowler, C. (1980). Coarticulation and theories of extrinsic timing control. *Journal of Phonetics*, 8, 113-133.

Fowler, C. A., Rubin, P., Remez, R. E., & Turvey, M. T. (1980). Implications for speech production of a general theory of action. In B. Butterworth, (Ed.), *Language production*. New York: Academic Press.

Fromkin, V. A. (Ed.) (1973). *Speech errors as linguistic evidence*. The Hague: Mouton & Co.

Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston: Houghton-Mifflin.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton-Mifflin.

Gleick, J. (1987). *Chaos: Making a new science*. New York: Viking.

Goldstein, L. M. (1983). Vowel shifts and articulatory-acoustic relations. In A. Cohen & M. P. R. van den Broeke (Eds.), *Abstracts of the Tenth International Congress of Phonetic Sciences* (pp. 267-273). Dordrecht: Foris Publications.

Haken, H. (1977). *Synergetics: An introduction*. Heidelberg: Springer Verlag.

Haken, H., Kelso, J. A. S., & Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. *Biological Cybernetics*, 51, 347-356.

Hofstadter, D. R. (1981). Strange attractors: Mathematical patterns delicately poised between order and chaos. *Scientific American*, 245, 22-43.

Keating, P. A. (1985). Universal phonetics and the organization of grammars. In V. A. Fromkin, (Ed.), *Phonetic linguistics*. New York: Academic Press.

Keating, P. A. (1988). The window model of coarticulation: articulatory evidence. *UCLA WPP, 69*,3-29.

Kelso, J. A. S. (1984). Phase transitions and critical behavior in human bimanual coordination. *American Journal of Physiology, 246*, R1000-R1004.

Kelso, J. A. S., & Scholz, J. P. (1985). Cooperative phenomena in biological motion. In (H. Haken (Ed.), *Complex systems: Operational approaches in neurobiology, physical systems and computers*. Berlin: Springer-Verlag.

Kelso, J. A. S., & Tuller, B. (1984). A dynamical basis for action systems. In M. S. Gazzaniga, (Ed.), *Handbook of cognitive neuroscience* (pp. 321-356). New York: Plenum.

Kelso, J. A. S., Vatikiotis-Bateson, E., Saltzman, E., & Kay, B. (1985). A qualitative dynamic analysis of reiterant speech production: Phase portraits, kinematics,and dynamic modeling. *Journal of the Acoustical Society of America, 77*, 266-280.

Kugler, P. N., & Turvey, M. T. (1987). *Information, natural law, and the self- assembly of rhythmic movement: Theoretical and experimental investigations*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Labov, W. (1972). The internal evolution of linguistic rules. In R. R. Stockwell & R. K. S. Macaulay (Eds.), *Linguistic change and generative theory* (pp. 101-171). Bloomington: Indiana University Press.

Labov, W., Yaeger, M., & Steiner, R. (1972). *A quantitative study of sound change in progress*. Philadelphia: The US Regional Survey.

Ladefoged, P. (1982). *A course in phonetics*. New York: Harcourt Brace Jovanovich.

Ladefoged, P. (1990). Some reflections on the IPA. *Journal of Phonetics, 18*, 335-346.

Lindblom, B., & Engstrand, O. (1989). In what sense is speech quantal? *Journal of Phonetics, 17*, 107-121.

Lindblom, B., MacNeilage, P., & Studdert-Kennedy, M. (1983). Self-organizing processes and the explanation of phonological universals. In (B. Butterworth, B. Comrie, & O. Dahl (Eds.), *Explanations of linguistic universals* (pp. 181-203). Mouton: The Hague.

Madore, B. F., & Freedman, W. L. (1987). Self-organizing structures. *American Scientist, 75*, 252-259.

Manuel, S. Y. (1990). The role of contrast in limiting vowel-to-vowel coarticulation in different languages. *Journal of the Acoustical Society of America, 88*, 1286-1298.

Manuel, S. Y., & Krakow, R. A. (1984). Universal and language particular aspects of vowel-to-vowel coarticulation. *Haskins Laboratories Status Report on Speech Research, SR-77/7*, 69-78.

May, R., & Oster, G. F. (1976). Bifurcations and dynamic complexity in simple ecological models. *The American Naturalist, 110*, 573.

McCarthy, J. J. (1988). Feature geometry and dependency: A review. *Phonetica, 45*, 84-108.

Mowrey, R. A., & MacKay, I. R. A. (1990). Phonological primitives: Electromyographic speech error evidence. *Journal of the Acoustical Society of America, 88*. 1299-1312.

Nearey, T. M. (1990). The segment as a unit of special perception. *Journal of Phonetics, 18*, 347-373.

Ohala, J. J. (1983). The origin of sound patterns in vocal tract constraints. In P. F. MacNeilage (Ed.), *The production of speech* (pp. 189-216). New York: Springer-Verlag.

Ostry, D. J., & Munhall, K. (1985). Control of rate and duration of speech movements. *Journal of the Acoustical Society of America, 77*, 640-648.

Pierrehumbert, J. (1990). Phonological and phonetic representation. *Journal of Phonetics, 18*, 375-394.

Prigogine, I., & Stengers, I. (1984). *Order out of chaos*. New York: Bantam Books.

Roberts, E. W. (1975). Speech errors as evidence for the reality of phonological units. *Lingua, 35*, 263-296.

Schmidt, R. C. (1988). *Dynamical constraints on the coordination of rhythmic limb movements between two people*. Doctoral dissertation. University of Connecticut.

Schmidt, R. C., Carello, C., & Turvey, M. T. (1990). Phase transitions and critical fluctuations in the visual coordination of rhythmic movements between people. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 227-247.

Schoner, G., & Kelso, J. A. S. (1988). Dynamic pattern generation in behavioral and neural systems. *Science, 239*, 1513-1520.

Shattuck-Huffnagel, S. (1983). Sublexical units and suprasegmental structure in speech production planning. In P. F. MacNeilage (Ed.), *The production of speech* (pp. 109-136). New York: Springer-Verlag.

Shattuck-Huffnagel, S. (1986). The representation of phonological information during speech production planning: Evidence from vowel errors in spontaneous speech. *Phonology Yearbook, 3*, 117-149.

Shattuck-Huffnagel, S. (1987). The role of word-onset consonants in speech production planning: New evidence from speech error patterns. In E. Keller & M. Gopnik (Eds.), *Motor and sensory processes of language* (pp. 17-51). Hillsdale, NJ: Lawrence Erlbaum Associates.

Shattuck-Huffnagel, S., & Klatt, D. (1979). The limited use of distinctive features and markedness in speech production: Evidence from speech error data. *Journal of Verbal Learning and Verbal Behavior, 18*, 41-55.

Stemberger, J. (1983). *Speech errors and theoretical phonology: A review*. Bloomington: Indiana University Linguistics Club.

Stevens, K. N. (1972). The quantal nature of speech: Evidence from articulatory-acoustic data. In E. E. David & P. B. Denes (Eds.), *Human communication: A unified view* (pp. 51-66). New York: McGraw-Hill.

Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics, 17*, 3-45.

Thompson, J. M. T., & Stewart, H. B. (1986). *Nonlinear dynamics and chaos*. New York: John Wiley & Sons.

Turvey, M. T. (1977). Preliminaries to a theory of action with reference to vision. In R. Shaw & J. Bransford (Eds.), *Perceiving, acting and knowing: Toward an ecological psychology*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Turvey, M. T. (1990). Coordination. *American Psychologist, 45*, 938-953.

Turvey, M. T., Rosenblum, L. D., Schmidt, R. C., & Kugler, P. N. (1986). Fluctuations and phase symmetry in coordinated rhythmic movements. *Journal of Experimental Psychology: Human Perception and Performance, 12*, 564-583.

Turvey, M. T., Saltzman, E., & Schmidt, R. C. (1991). Dynamics and task-specific coordinations. In N. I. Badler, B. A. Barsky, & D. Zeltzer (Eds.), *Making them move* (pp. 157-170). Mountain View, CA: Morgan Kaufmann.

Vatikiotis-Bateson, E. (1988). *Linguistic structure and articulatory dynamics*. Bloomington: Indiana University Linguistics Club.

Vitz, P. C., & Winkler, B. S. (1973). Predicting the judged "similarity of sound"of English words. *Journal of Verbal Learning and Verbal Behavior, 12*, 373- 388.

Wood, S. (1982). X-ray and model studies of vowel articulation. *Working Papers, Lund University*, 23.

FOOTNOTES

[*] *Journal of Phonetics, 18*, 411-424 (1990).

[†] Also Yale University Department of Linguistics.

Haskins Laboratories Status Report on Speech Research
1991, SR-105/106, 93-106

Young Infants' Perception of Liquid Coarticulatory Influences on Following Stop Consonants*

Carol A. Fowler,† Catherine T. Best,†† and Gerald W. McRoberts†††

Phonetic segments are coarticulated in speech. Accordingly, the articulatory and acoustic properties of the speech signal during the time-frame traditionally identified with a given phoneme are highly context-sensitive. For example, due to carryover coarticulation, the front tongue-tip position for /l/ results in more fronted tongue-body contact for a /g/ preceded by /l/ than for a /g/ preceded by /r/. Perception by mature listeners shows a complementary sensitivity—when a synthetic /da/-/ga/ continuum is preceded by either /al/ or /ar/, adults hear more /g/s following /l/ than /r/. That is, some of the fronting information in the temporal domain of the stop is perceptually attributed to /l/ (Mann, 1980). We replicated this finding and extended it to a signal-detection test of discrimination with adults, using triads of disyllables. Three equidistant items from a /da/-/ga/ continuum were used preceded by /al/ and /ar/. In the identification test, adults had identified item ga5 as "ga," and da1 as "da," following both /al/ and /ar/, whereas they identified the crucial item d/ga3 predominantly as "ga" after /al/ but as "da" after /ar/. In the discrimination test, they discriminated d/ga3 from da1 preceded by /al/ but not /ar/; compatibly, they discriminated d/ga3 readily from ga5 preceded by /ar/ but poorly preceded by /al/. We obtained similar results with four month old infants. Following habituation to either ald/ga3 or ard/ga3, infants heard either the corresponding ga5 or da1 disyllable. As predicted, the infants discriminated d/ga3 from da1 following /al/ but not /ar/; conversely, they discriminated d/ga3 from ga5 following /ar/ but not /al/. The results suggest that prelinguistic infants disentangle consonant-consonant coarticulatory influences in speech in an adult-like fashion.

The mappings are complex between the phonetic structure of a spoken message and the acoustic structure in the speech signal that conveys the message to a listener. So too, therefore, is the reverse mapping between acoustic signal and phonetic message. Of course, mature listeners recover phonetic properties despite the complexity of these mappings. Adults have extensive experience hearing and producing the sounds of speech, as well as an active knowledge of the lexicon and syntax of their language, all of which potentially aid recovery of a speech message. Yet what of very young infants, who have much more limited speech listening experience, even less experience producing speechlike sounds, and no comprehension of words or syntactic rules? What structure do they recover from the acoustic speech signal? Certainly, the acquisition of language entails recovery of phonetic structure from the acoustic signal. But when does the capability to recover phonetic structure emerge? Previous findings indicate that certain relevant achievements, such as perceptual constancy, perceptual equivalence and trading of phonetically-equivalent acoustic properties, and use of context in speech perception, are present long before the infant utters or understands its first meaningful word,

This research was supported by NIH grant DC00403 to the second author. We wish to thank the following people for their contributions to completion of the project: Virginia Mann for a helpful discussion of a possible auditory account of the results; Michael Donaghu for help collecting and scoring the data of the adult listeners in Experiments 1 and 2; and Glendessa Insabella, Stephen Luke, Peter Kim, Laura Klatt, Meredith Russell, Jean Silver, Pam Speigel, and Jane Womer for help collecting, scoring and analyzing the infant data in Experiment 3. We also thank our adult subjects, and are particularly grateful to the parents of the infant subjects for their interest in the project and their willingness to permit their children's participation.

and even before it begins to produce syllable-like babbles (Bertoncini, Bijeljac-Babic, Jusczyk, Kennedy, & Mehler, 1988; Eimas, 1985; Eimas & Miller, 1980a, 1980b; Grieser & Kuhl, 1989; Kuhl, 1979, 1980, 1983; Morse, Eilers, & Gavin, 1982).

None of those reports, however, has focused on infants' perception of the particular complex mappings between acoustic and phonetic structure that arise from coarticulation. Coarticulation is of particular interest because of the ways in which it complicates the acoustic consequences of phonetic-segment production. The language-learning child must disentangle those complications in order to come to recognize the segmental structure of speech.

Talkers coarticulate phonetic segments—that is, they implement the phonetic properties of neighboring consonants and vowels in overlapping time frames. The effects work in both directions in time. As an example of anticipatory coarticulation, vowels followed by nasal consonants are nasalized (e.g., Kent, Carney, & Severeid, 1974); as an example of carryover, or perseverative, coarticulation, /g/ preceded by /l/ is fronted (Mann, 1980). The consequence of such coarticulatory overlap is that coarticulating phonetic segments have converging effects on common acoustic dimensions of a speech signal within a given time frame (e.g., Fant & Lindblom, 1961). Accordingly, one must ask how even mature listeners deal with the converging effects of diverse segmental properties on common acoustic dimensions. Research shows that adults deal remarkably successfully with the convergences, behaving as though they have disentangled the converging influences on the acoustic signal. Listeners treat acoustic information for a segment *x*, occurring in the temporal domain of segment *y*, as information for *x*. This holds, for example, for anticipatory vowel information that appears in the domain of a preceding fricative (Whalen, 1983) or in the domain of an earlier transconsonantal vowel (Fowler & Smith, 1986; Martin & Bunnell, 1981); it also holds for anticipatory information about a nasal consonant that appears in the temporal domain of a preceding vowel (Krakow, Beddor, Goldstein, & Fowler, 1988), and for the carryover effects of one consonant occurring in the domain of another (Mann, 1980). In the last-cited research, the high front (alveolar) position of tongue-tip contact for an /l/ pulls the tongue-body forward, whereas /r/ does not exert a fronting effect. As a result, the velar contact for a /g/ is pulled forward in the mouth (i.e., F3 onset frequency is raised in the direction of the F3 onset frequency for /da/) when it is preceded by an /l/ but not when preceded by an /r/. Compatible with this, if a synthetic continuum from /da/ to /ga/ is preceded by either /al/ or /ar/, adults hear more /g/s following /l/ than /r/, indicating that some of the tongue-fronting information that occurs in the temporal domain of the stop consonant is perceptually attributed to the preceding /l/ (Mann, 1980).

In addition to the classic coarticulatory effects just described, prosodic and nonlinguistic properties of an utterance are coproduced with phonetic segments, and converge with the segmental influences on the acoustic signal. For example, prosody affects the durational properties and fundamental frequency (F_0) of an utterance, both of which also reflect systematic variation due to the consonants and vowels on which the prosody is realized (e.g., Klatt, 1976; Silverman, 1987). Rate variation illustrates nonlinguistic influences. In speaking, changes in rate have durational effects that may converge with phonetic variation (for example, durational differences related to vowel height), phonological-segmental variation (e.g., differences in phonological length) and prosodic variation (e.g., durational differences related to stress patterns). As in cases of segmental coarticulatory influences, listeners apparently disentangle the prosodic and nonlinguistic influences on the signal. For example, they judge intonational accents as if the effects of vowel height on F_0 had been eliminated (Silverman, 1987), while, for its part, the contribution of vowel height to the F_0 contour is used as information for vowel height (Reinholt-Peterson, 1986). In addition, the effects of speech rate variations are effectively eliminated from the phonetic sources of variation in formant-transition duration that distinguish /b/ from /w/ (e.g., Miller & Liberman, 1979).

The question arises whether the ability to perceive phonetic segments with these converging influences disentangled requires experience producing coarticulated speech. That is, must the speaker/hearer learn to associate the intended phonetic segments with their complex and temporally overlapping acoustic consequences? The prebabbling infant under about 7 months of age lacks this kind of experience because it is not yet producing syllabic combinations of consonant-like and vowel-like sounds. The relevant articulatory experience might be acquired, then, during the last half of the first year, as the infant begins to produce reduplicated and nonreduplicated babbling (e.g., Oller, 1980; Stark, 1980). Alternatively, the relevant factor may not be articulatory experience *per se*, but rather the development of a sizable lex-

icon beyond 50 or so words, which may enable the child to recognize the efficiency of using a phonological system for lexical organization. We suspected, however, that adult-like perceptual disentangling of coarticulatory influences in the speech signal might be evident even earlier in development than either of these possibilities. Our prediction was derived from an account of speech perception that posits articulatory gestures as the primitives of both speech perception and speech production (Best, in press; Fowler & Rosenblum, 1990; see also Liberman & Mattingly, 1985). The specific reasoning that led to the studies reported here was that young infants should show perceptual sensitivity to coarticulatory influences as a consequence of a basic perceptual tendency to recover information in stimulation about the source event that produced the signal (e.g., Gibson, 1966, 1979). To test our hypothesis, the present study examined how very young, pre-babbling infants handle coarticulatory influences when perceiving speech. Findings on this issue are also relevant to accounts that focus on basic auditory processes (e.g., Diehl & Kluender, 1989); we address two such accounts in our General Discussion.

Infants do show evidence, in other domains, of adult-like perception of the acoustic speech signal. For example, they exhibit perceptual equivalence of temporal and spectral information for a stop consonant in a "say"-"stay" context (Eimas, 1985; see also Morse et al., 1982; cf. Eilers & Oller, 1989).[1] This pattern replicates earlier findings with adults by Best et al. (1981; see also Fitch, Halwes, Erickson, & Liberman, 1980; review by Repp, 1982). Infants also show shifts in boundaries between voicing categories along a voice-onset time (VOT) continuum as the starting frequency of F_1 is varied, demonstrating a trading relation between temporal and spectral information about stop voicing (Miller & Eimas, 1983), again in keeping with adult findings (Summerfield & Haggard, 1977). Finally, as Carden, Levitt, Jusczyk, and Walley (1981) had found earlier in a study of context effects in adult speech perception, infants fail to distinguish fricationless /fa/ and /θa/, but do distinguish them when the same frication noise is placed before the truncated syllables (Levitt, Jusczyk, Murray, & Carden, 1989).

Specifically regarding infants' handling of the convergence of multiple aspects of linguistic structure on a single acoustic dimension, however, less is known. They do show adult-like normalization for the influence of a nonlinguistic factor—speech rate variations—when discriminating /b/-/w/ syllables that vary in formant-transition duration (Miller & Eimas, 1983; cf. Jusczyk, Pisoni, Reed, Fernald, & Myers, 1983). To our knowledge, however, no one has looked at infants' perception of convergences caused by concurrent production of multiple linguistic properties of an utterance, in particular, coarticulation of segmental properties. As we suggested earlier, perceptual disentangling of the acoustic effects of multiple gestural influences on the speech signal are important to the child's discovery of the segmental organization of its native language.

Therefore, in the present study we examined prelinguistic infants' ability to separate coarticulatory influences on a speech signal, before the age at which infants begin to produce syllabic babbling themselves. We chose to use Mann's (1980) stimuli,[2] because experience producing /r/ and /l/, and CC sequences in general, typically emerges rather late in language development, during the preschool years; those properties are not evident in the vocalizations of 4-5 month olds, and are rare even in the babbling of much older infants. The first two experiments with adult listeners were designed to verify earlier findings of perceptual "normalization" of coarticulatory influences between adjacent consonants, and to extend those findings to performance under conditions similar to those used in infant discrimination testing procedures. These first two studies also served to identify the appropriate stimulus pairings for use in the final experiment with 4-5 month old infant listeners. We predicted that, even prior to producing syllable-like babbling, infants would show the same pattern of perceptual sensitivity to coarticulatory influences as adults.

EXPERIMENT 1

The first experiment replicated a portion of Mann's (1980) Experiment 1 using a subset of her stimuli. In Mann's research on adult listeners, the boundary along a synthetic /da/ to /ga/ continuum was shifted by a preceding naturally-produced /al/ syllable as compared to a preceding /ar/ or no preceding syllable at all. Specifically, /ga/ responses increased in the context of /al/. Mann interpreted the findings as suggestive evidence that perception takes into account the carryover coarticulatory fronting effects of /l/ on a following velar consonant when identifying a following consonant as having a velar or alveolar place of articulation. Our primary purpose in this study was to determine whether we could identify the critical stimulus items needed for the infant test (Experiment 3) and for an adult test under conditions approximating those of the infant

discrimination procedure (Experiment 2). Specifically, the latter two procedures required that we obtain three equidistant items along the /da/-/ga/ continuum, one of which adults identify consistently as /da/ in both the /al/ and the /ar/ context, one consistently identified as /ga/ in both contexts, and a crucial item midway between these two which is identified predominantly as /ga/ following /al/ but as /da/ following /ar/.

Methods

Subjects

Subjects were nine undergraduate students and one graduate student. All were native speakers of English who reported normal hearing, and all were naive to the purposes of the experiment. Undergraduates received course credit for their participation.[3]

Materials

We used a subset of Mann's stimuli. They consisted of "hybrid" disyllables of which the first syllable was naturally-produced and the second was synthesized. Use of natural initial syllables ensures that natural coarticulatory information for a following stop consonant is available to the listeners; use of synthetic final CV syllables permits sensitive detection of shifts in identification of the synthetic consonant along a continuum according to coarticulatory context.

The first syllables of each disyllabic nonsense word were stressed /al/ or /ar/ produced by a male speaker of English in the context of following /da/ or /ga/. Durations of each of the four precursor syllables were as follows: "al(da)" 261 ms, "al(ga)" 262 ms, "ar(da)" 248 ms and "ar(ga)" 242 ms. As Mann's (1980) measurements indicate, major differences between /al/ and /ar/ syllables are that /ar/ has a higher F2 and a lower F3 than /al/. For the four syllables we used, estimates of the offset frequencies of F2 and F3, were, respectively, 1012 and 2720 Hz for "al(d)," 1060 and 2720 Hz for "al(g)," 1566 and 1824 for "ar(d)" and 1402 and 2018 Hz for "ar(g)." In the isolated /ar/ and /al/, the place of articulation of the stop consonant following the /r/ or /l/ in the original disyllabic productions was identifiable due to anticipatory coarticulation. Each /al/ and /ar/ syllable was spliced onto each member of a seven item /da/ to /ga/ synthetic speech continuum to create four distinct VCCV continua. Stimuli in the CV synthetic continuum differed in the onset of F3, which ranged from 2690 to 2104 Hz in approximately even steps. Onsets of F_1 and F_2 were 310 and 1588 Hz . Steady states for F_1, F_2 and F_3 were 649, 1131 and 2448 Hz. Transitions were 100 ms in duration. While these are rather long transitions for stop consonants, we chose to retain Mann's original stimuli; in any case, they were clearly stops rather than glides. Total CV durations were 230 ms including a 50 ms closure interval following the /al/ or /ar/ precursor.

Pairing of each natural VC syllable with each continuum member gave 28 distinct disyllables. A test order was created consisting of 10 tokens of each of the 28 disyllables in random order with 3.5 seconds between trials in the test and a seven second pause after each block of 28 stimuli.

Procedure

Subjects listened to tape-recorded stimulus presentations over headphones in a sound-attenuated room. They were tested in groups of 1-3 students. They were instructed to identify the second consonant in each disyllable as "d" or "g" (by writing the appropriate letter on an answer sheet), guessing if necessary.

Results and Discussion

Figure 1 displays the percentage of "g" responses to synthetic CV continuum members separately for the four continua. The top display in the figure compares the outcome when precursor syllables are "al(d)" and "ar(d)"; the bottom display presents the results when precursors are "al(g)" and "ar(g)." In an analysis of variance with factors continuum (items 1-7), precursor syllable (/al/ or /ar/) and stop context of the precursor as originally produced (/d/ or /g/), all main effects and interactions reached significance. The main effect of continuum ($F(6,54) = 144.76$), which accounted for most of the variance in the analysis (72%), reflected the increase in "g" responses with an decrease in onset F3 in the synthetic continuum. The main effect of precursor ($F(1,9) = 29.33$), $p = .0005$) reflected the effect of interest, a lower percentage of "g" responses associated with the precursor /ar/ as compared to /al/. The main effect of contextual stop ($F(1,9) = 6.43$, $p = .03$) reflected a lower percentage of "g" responses for precursors originally produced in the context of following /d/ than /g/. Interactions involving the factor continuum appeared largely to reflect the smaller magnitude of main effects and interactions at the endpoints of the continuum where "g" responses were at floor or ceiling. The interaction of precursor syllable × context stop consonant ($F(1,9) = 14.04$, $p = .0046$) was significant because the effect of context consonant was present only for the /ar/ precursor, and, on the other side, because the effect of precursor syllable was present only for the "al(d)-ar(d)" precursor pair. Mann (1980) obtained this

interaction as well (see her Figure 3); however, her effect of precursor syllable was reduced, rather than eliminated, for the "al(g)-ar(g)" precursors.

Just one of the two possible pairs of continua that we might use with infants provided an outcome meeting our requirements. With precursors "al(d)" and "ar(d)", as depicted in Figure 1 (top) the fifth CV along the continuum (henceforth ga5) was identif 1 predominantly as "ga" preceded by both precursor syllables (97% of the time after /al/ and 72% after /ar/), while the first (da1) was identified predominantly as "da" in both contexts (93% after /al/ and 98% after /ar/). The crucial third CV (henceforth d/ga3) was identified predominantly as "ga" after /al/ (70%), but as "da" after /ar/ (90%). Pairing these CVs with /al/ and /ar/ allowed us to test two between-category discriminations in Experiments 2 and 3, one for each preceding context (ald/ga3 versus alda1 and ard/ga3 versus arga5), and two within-category discriminations (ald/ga3 versus alga5 and ard/ga3 versus arda1), with the acoustic differences matched among between- and within-category pairs. Thus, the within- and between category pairs pattern oppositely between the /al/ context and the /ar/ context.

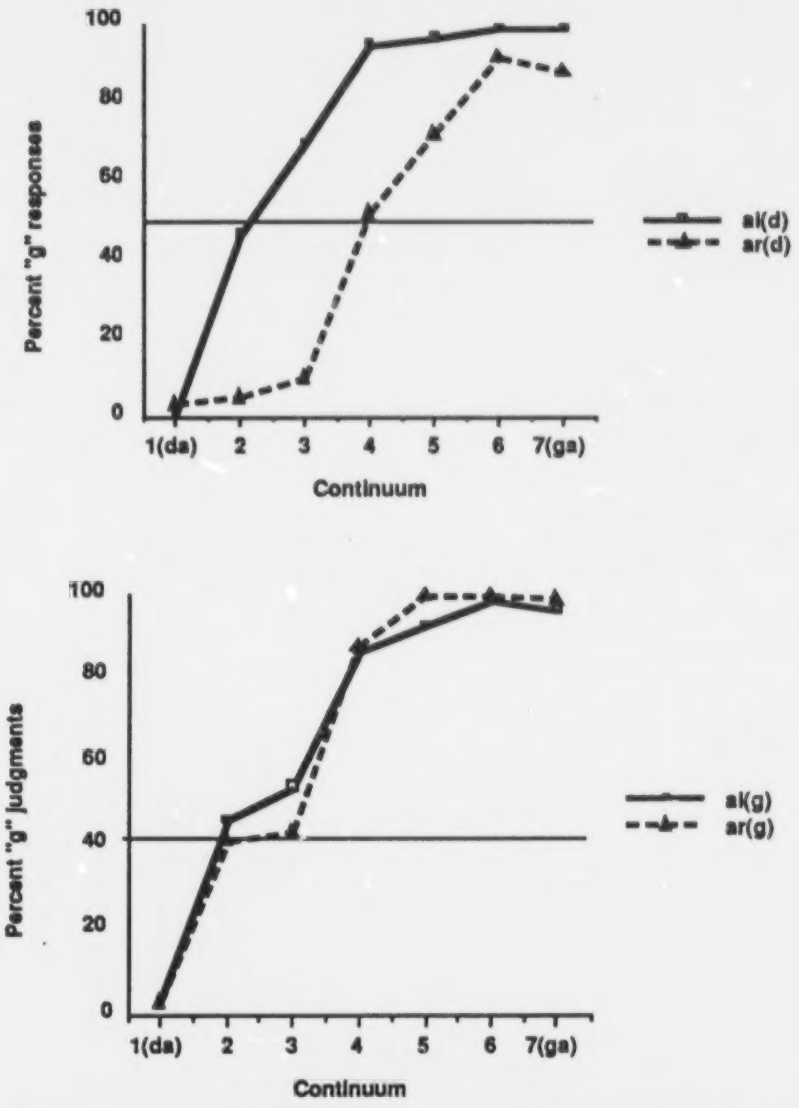


Figure 1. **Identification functions averaged across 10 adult listeners, for synthetic /da/-/ga/ continuum preceded by stressed "al(d)" and "ar(d)" (top) in Experiment 1. Bottom: data on "al(g)" and "ar(g)" continua.**

In the other possible pair of continua (with "al(g)" and "ar(g)" precursors; Figure 1 bottom), while continuum members 5 and 1 were convincingly "ga" and "da" respectively (with percent identification >92% in each response category), and while responses to the third continuum member were predominantly "ga" with the "al" precursor (55%) and "da" with the "ar" precursor (56%), the 11% separation in response rates to the third continuum member was small and unreliable (t(9) = 1.03). Possibly the precursor effect diminishes (Mann, 1980) or, here, is eliminated, in the context of following /g/ because information for /g/ in "ar(g)" promotes "ga" identifications more so than does /g/ information in "al(g)."[4] This effect of anticipatory coarticulation on "g" identifications in the "ar(g)" context balances the complementary effect of carryover coarticulation on listeners' tendency to report more "g"s following "al" than "ar" in the "ar(g)-al(g)" continua. As for reasons why effects of the precursor syllables originally followed by /g/ were present in Mann's findings and not in our own, the most likely reason is that we used just one of her six (three stressed and three unstressed) tokens of each precursor syllable. Rather than pursue this issue, however, which was not a primary focus of our study, we dropped the "al(g)" and "ar(g)" precursors and performed the remaining experiments with "al(d)" and "ar(d)" precursors.

Testing the foregoing between- and within-category discriminations using "al(d)" and "ar(d)" precursors with prelinguistic infant listeners may help to determine whether pre-babbling infants show an adult-like effect of precursor syllable on their responses to continuum members. Before testing infants, however, we ran a further study with adults. Experiment 2 was designed to ensure that adult discrimination performance, under conditions similar to the infant discrimination procedure used in Experiment 3, would reflect the categorizations suggested by the identification data collected in Experiment 1.

EXPERIMENT 2

For Experiment 2 we chose a signal-detection discrimination procedure for adults. This was necessary to verify that the stimulus pairs we had chosen based on the results of Experiment 1 would maintain their category memberships when presented under listening conditions that approximated the discrimination task we planned to use with our infant listeners. Accordingly, adults listened to sequences of varying numbers of identical (background) disyllables (either of the critical stimuli ald/ga3 or ard/ga3), in which a new disyllable (/al/ or /ar/ followed by either da1 or ga5) was presented at an unpredictable point near the end of the sequence. They hit a response key whenever they detected a change from the background disyllables. We performed a signal detection analysis on the data.

Methods

Subjects

Subjects were 12 undergraduates who participated for course credit. All were native speakers of English who reported normal hearing. All were naive to the experimental hypotheses.

Materials

The test consisted of 48 sequences evenly divided among the four conditions of the experiment (background disyllable ald/ga3 changing either to alda1 or alga5 and analogous sequences using ard/ga3 changing either to arda1 or arga5). Across sequences, the change or target disyllable occurred after as few as 10 repetitions of the background disyllable or as many as 33 repetitions. The target disyllable was presented one time in each sequence and it was followed by two repetitions of the background disyllable before the sequence ended. Distance of the target disyllable from the beginning of the sequence was balanced across lists. There was a 1500 ms interval (offset to onset) between disyllables in a sequence. On the second channel of the tape, a tone pulse marked the onset of each disyllable. That pulse, input to a computer, enabled association of key press responses signaling detection of a target disyllable with each disyllable in a sequence.

Procedure

Listeners were tested individually. Stimuli were presented over a loudspeaker (as in the infant experiment) in a quiet listening room. Subjects were instructed to hit a key on a computer terminal keyboard whenever they heard a change from the background disyllable, however subtle the change might be. They were not told that there was just one target disyllable per sequence; accordingly they were allowed to hit the key as many times as they chose on each trial of the experiment. They were told, however, that the change would never occur before the eleventh disyllable of a given trial; this would allow them to get used to the background disyllable's sound before listening for a change.

Measures were hits, misses, false alarms and correct rejections, converted to d' measures.

Results

Figure 2 displays the *d'* s for the four conditions. As the figure shows, *d'* measures were considerably higher for the two between-category discriminations than for their corresponding within-category discriminations. In an analysis of variance with repeated measures factors "precursor syllable" (/al/ or /ar/) and "direction of shift" (to ga5 or da1), neither main effect was significant (both Fs< 1), but the interaction was highly significant ($F(1,11) = 57.77$, $p < .0001$). The interaction reflects two significant outcomes: 1) poor discrimination (*d'* = .07) of d/ga3 from da1 in the context of /ar/, but good discrimination of the same shift in the context of /al/ (*d'* =2.38) and 2) poor discrimination of d/ga3 from ga5 in the context of /al/ (*d'* =.57), but good discrimination in the context of /ar/ (*d'* =2.40). Pairwise comparisons (Scheffé tests) verified that *d'* s for the between-category discriminations are significantly larger than those for within-category discriminations (/al/: $F(1,11) = 7.32$, $p = .006$; /ar/: $F(1,11) = 12.14$, $p = .0009$). Pairwise comparisons of the two between-category discriminations and of the two within-category discriminations were nonsignificant (both Fs < 1). Finally, excepting the *d'* values for the within-category discrimination with /ar/ as the precursor syllable, all conditions show significantly positive *d'*s, indicating significant evidence of discrimination (for the within- category discrimination involving /al/: $t(11) = 2.97$, $p = .01$; there were no negative *d'*s in the two between-category conditions.

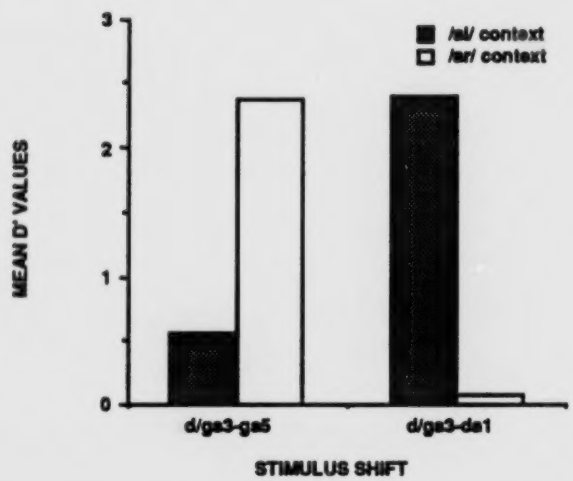


Figure 2. **Average d' values of 12 adult listeners in the signal detection test for discrimination of d/ga3 from ga5 and from da1 preceded by /al/ and by /ar/ (Experiment 2).**

On the basis of these findings we considered our stimulus pairings appropriate for testing with prelinguistic infants.

EXPERIMENT 3

In the final experiment, we examined 4-5 month olds to determine whether or not they disentangle coarticulatory influences as adults do. We predicted that our pre-babbling infants would discriminate the stimulus pairs determined to be between-category in the adult tests, but would fail to discriminate the pairs that were within-category for adults. That is, the infants should show the same context-dependent reversal in performance levels as had the adults in Experiment 2 when discriminating d/ga3 from ga5 and from da1, suggesting perceptual sensitivity to the converging influences of multiple phonetic segments on a single acoustic dimension.

The infants participated in a habituation procedure comparable to the signal detection task of the adults in Experiment 2. Following habituation to either the ald/ga3 or ard/ga3 disyllable, infants received one of two stimulus shifts: to the corresponding ga5 disyllable or the corresponding da1 disyllable. Fixation time before and after the shift was examined for evidence of dishabituation to the novel stimuli.

Methods

Subjects

Subjects were 48 infants from the communities surrounding Wesleyan University, between 4 and 6 months of age (M = 4 months, 17 days; range = 4 months to 5 months, 29 days). Twelve infants were tested in each of the four test conditions (see Procedure), with males and females approximately equally distributed across conditions. Data from an additional 16 infants were excluded because of crying/fussing (3), inattention to the visual stimulus (6), performance scores greater than 2 S.D. beyond the mean for the infant's test condition (1), equipment problems (2), and experimental error (4). Thus, the success rate was 75%. The drop-out rate was approximately evenly distributed across the experimental conditions.

The subjects were solicited via mailings and follow-up phone calls to parents listed in the birth announcements of newspapers for Middletown CT and neighboring towns. This recruitment procedure yields an approximate 25-30% acceptance rate.

Materials

There were four 30-minute stimulus tapes, one for each test condition. Stimuli were recorded in

synchrony on two channels of a four-track tape, with tone pulses recorded on a third track, 15 ms preceding the onsets of each pair of items on the stimulus channels. There were 1500 ms interstimulus intervals between disyllables on the stimulus channels, as in Experiment 2. The tone pulses were used to signal a computer as to when stimulus presentations could be initiated, terminated, or switched between channels (see Procedure). The d/ga3 stimulus preceded by the precursor syllable for the appropriate condition (/al/ or /ar/) was recorded on one channel of the tape, while synchronized repetitions of the appropriate ga5 or da1 disyllable were recorded on the other channel.

Procedure

Each subject was tested on one of the four test comparisons: 1) ald/ga3 → alga5; 2) ald/ga3 → alda1; 3) ard/ga3 → arga5; 4) ard/ga3 → arda1. Conditions 1 and 4 presented within-category comparisons according the the adult findings, whereas conditions 2 and 3 presented between-category comparisons.

We employed the infant-controlled visual fixation discrimination procedure described by Miller (1983). In this procedure, the infant is operantly conditioned to fixate a rear-projected slide of a brightly-colored checkerboard in order to receive audio presentations of speech stimuli. The stimuli were presented at a comfortable listening level (70 db) over a loudspeaker (Jamo) hidden a few feet above the target slide. A computer (Atari-800) initiated and terminated the stimulus presentations from a continuously-playing tape deck (Otari 5050 MXB), and determined which channel of the tape was presented over the loudspeaker, based on key-press input from a trained observer. The observer viewed a videomonitor conveying input from a hidden camera focused on the infant's face (under control of a cameraperson) in order to detect the infant's fixations of the target slide. The observer was separated from the infant and loudspeaker by a sound-treated wall. To further assure that (s)he was "deaf" to the stimuli that the infant heard, the observer wore headphones and listened to music throughout the session. In addition, the observer was unaware of when during the test the stimulus shift trials actually occurred, because the number of habituation trials varied from infant to infant depending on their fixation patterns. The observers' lack of awareness about the course of the test session was underscored by the fact that the cameraperson invariably had to let them know when the test had ended.

The infant's fixation behavior determined the division of the test session into individual trials. Whenever the infant gazed away from the target slide for more than 2 sec, the slide was automatically shut off for 1 sec and then redisplayed to begin a new trial. Once the infant habituated to the familiarization stimulus during the habituation phase of the test, the speech presentations were shifted to the novel stimulus on the second audio channel during the test phase. The habituation criterion was a decline in the infant's fixations on two consecutive trials to a level below 50% of the mean of the two highest preceding trials. Stimulus presentations were shifted to the test channel on the next trial following that on which the habituation criterion was met. The exact details of the procedure and experimental set-up are described in Best, McRoberts, and Sithole (1988).

To assess the inter-judge reliability of observations of the infants' visual fixations, the videotapes of 29 test sessions were re-scored by members of the research team (60% of the sessions). Included were all sessions for which there was any question about the infants' fixation pattern and/or behavioral state (e.g., fussing), as well as an equal number of unquestioned sessions. Inter-observer correlations were quite high, ranging between .95 and .99, with one exception at .78 (the latter session was retained because the single test trial on which the observers disagreed was not one of the critical trials surrounding the stimulus shift).

Results and Discussion

We computed the mean looking times for the two trials immediately preceding the stimulus shift (habituation level) and for the first two postshift trials beginning when the infant heard at least one test stimulus presentation (dishabituation). Some infants failed to look at the slide during the first trial or so after the shift because they had habituated to 0 during the first part of the test, and hence they failed to hear any postshift stimuli during those first postshift trials. Because at least one postshift stimulus was needed for the infant to have an opportunity to discriminate between preshift and postshift stimuli, then, we did not include in the dishabituation mean any non-looking trial(s) immediately following the shift. Once the infant looked even briefly enough to hear one postshift stimulus, the true dishabituation trials began (see Best et al., 1988). The summary data are shown in Figure 3 for the four conditions of the experiment. Qualitatively, the response pat-

tern in Figure 3 is very similar to that of the adult listeners shown in Figure 2. As predicted, *t* tests (one-tailed) revealed significant recovery after the stimulus shift in the two conditions predicted to provide between-category comparisons (ald/ga3 to alda1: $t(11) = 2.74$. $p = .01$ and ard/ga3 to arga5: $t(11) = 2.04$, $p = .03$), and no significant recovery in the remaining conditions. Compatibly, an analysis of variance on pre- and post-habituation looking times with factors "precursor syllable" (/al/ or /ar/) and "direction of shift" (to ga5 or to da1) yielded no main effects (both Fs < 1) but a significant interaction ($F(1,44) = 4.57$, $p = .038$). The interaction is significant because the relative recovery magnitudes in the two shift directions (ga5, da1) pattern oppositely depending on the preceding context.

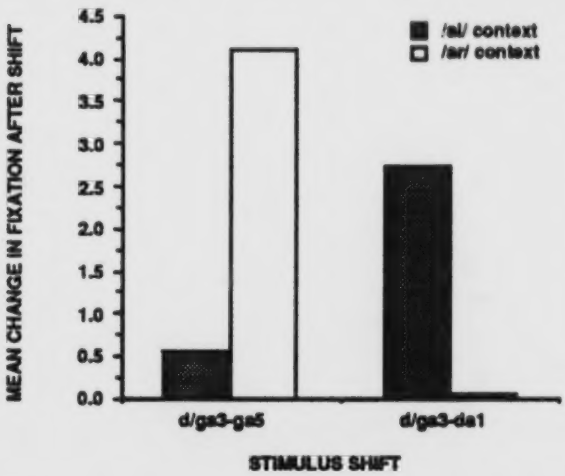


Figure 3. Infants' response recoveries (in sec) following the stimulus change in each condition (12 Ss per condition) of the infant-controlled visual fixation habituation procedure; results indicate extent of infant discrimination of d/ga3 from ga5 and from da1 preceded by /al/ or /ar/ (Experiment 3).

Accordingly, for prelinguistic infants as for adults, a stop consonant that is ambiguous between /d/ and /g/ is heard as less "d"-like in the context of /l/ than in the context of /r/. That is, both mature listeners and prelinguistic infants effectively remove the coarticulatory fronting influence that /l/ has on a following velar consonant.

GENERAL DISCUSSION

That prelinguistic infants show the same interaction in the two syllable contexts as adults do demonstrates conclusively that, in this instance at least, neither experience producing coarticulated speech, nor acquisition of language-specific lexical items is required for perceptual elimination of coarticulatory influences on acoustic information for a phonetic segment. Another finding in the literature is relevant to an interpretation of the outcome. Mann (1986) tested Japanese listeners on the disyllables used in Mann (1980) and in the present experiments. This language group is of interest because the Japanese language does not make a phonemic /l/-/r/ distinction. Mann identified two groups of Japanese listeners based on their ability to label stimuli consistently as "l" and "r." In one group, listeners were at chance on average (58% correct, $p > .1$) in identifying the final consonants of /al/ and /ar/. In another, they were near perfect (98% correct). Remarkably, both groups of listeners showed shifts in the /da/-/ga/ boundary in the context of preceding /l/ as compared to /r/. Moreover, the magnitude of the shift was the same in the two groups of Japanese listeners as in a third group of native English listeners. Apparently a listener need not be able to classify consonants into distinct phonological categories in order to extract their different coarticulatory influences on neighboring consonants. How, then, is the extraction to be explained?

If both mature listeners who cannot reliably classify /l/s and /r/s into different phonemic categories and prelinguistic infants show the same perceptual response patterns as mature listeners who command the phonemic distinction, presumably an explanation for the response patterns must derive from something that all three groups have in common. One possibility is the auditory systems of these listeners.

Mann (1986) considers and rejects one such account of the Japanese listeners' performance patterns. It is that auditory nerve fibers are known to exhibit forward masking by one acoustic signal that precedes another by 50-100 ms. The masking effect is such that the response of the auditory nerve is depressed to stimuli in the same frequency range as the preceding masking stimulus (Delgutte & Kiang, 1984; Harris & Dallos, 1979; Smith, 1977). Psychophysical tests of human listeners reveal compatible response patterns (Elliot, 1971; Moore, 1978).

In Mann's stimuli /al/, but not /ar/ has an F_3 offset frequency close to the onset frequency of F_3 for stimuli at the /da/ end of the /da/-/ga/ continuum. Accordingly, preceding /al/ should selectively depress auditory-nerve sensitivity to stimuli at that end of the continuum, giving rise to the observed increase in "ga" responses.

For several reasons, we reject this account of our findings and of Mann's (1980; 1986). First, as Mann (1986) points out, the auditory masking interpretation is weakened by findings of Mann and Liberman (1983). That study employed the same stimuli under test here; however, the critical F_3 transitions for /da/ or /ga/ were presented to one ear and the remainder (base) of the disyllable was presented to the other ear. This manner of presenting speech stimuli gives rise to a "duplex" percept in which the F_3 transition is apparently heard in two ways at once. It is integrated with the information in the opposite ear, giving rise, in that location, to a /da/ or /ga/ percept for the second syllable of the disyllable; it is simultaneously heard as a pitch glide in the ear receiving the transition. Under these conditions, Mann and Liberman obtained two findings that are important for the present purposes. First, context effects of /l/ on "d" and "g" classifications were present, eliminating the auditory nerve (or in fact any other peripheral influence) as a source of the context effects. Second, context effects were absent in the classifications of the pitch glides, weakening any account of the context effects that ascribed them to masking originating in higher-level (central) auditory-system processing *per se*.

A final reason to reject an auditory masking account is that the offset frequency of F_3 of /l/ (2711 Hz averaged across the multiple natural /al/ tokens in Mann's stimuli) is closest to the endpoint /da/'s F_3 onset frequency (2690 Hz) and becomes progressively farther from the other continuum members' F_3 onsets as we approach ga7 (2104 Hz). Since, in the auditory masking literature, effects are largest for stimuli closest in frequency to the context stimulus, auditory effects should be largest on the /da/ endpoint and progressively smaller thereafter (Mann, personal communication). However, this is opposite to the pattern of context effects found in Mann (1980, 1986), Mann and Liberman (1983), and the present study. Further, masking should be absent outside the critical band surrounding 2711 Hz (approximately 400 Hz), but the first continuum member outside that band is d/ga3, the stimulus on which the largest context effects were obtained.

If the perceptual elimination of coarticulatory influences is not to be explained by appeal to masking, how is it to be explained? Possibly the findings of Mann and Liberman permit a further inference about the domain in which an explanation for the context effects should be sought. Mann and Liberman found that only formant transitions that are experienced in the same spatial location (ear) as the rest of the disyllable and that are experienced as part of the disyllable are subject to context effects. The dichotic shift in perceived location of the transition must be associated with a perceptual "parsing" of the acoustic signal in which the transitions and the remainder of the disyllable serve as joint acoustic consequences of a single coherent sound-producing event. If so, then context effects may arise only when the context counts perceptually as part of the same sound-producing event that gave rise to the transitions. Yet parsing into distinct segmental influences on a single sound-producing event must be based on relevant information in the acoustic signal. If so, perhaps there is also an informational basis in the signal for the context effects, rather than a basis in the auditory mechanisms of the listener.

Consider one implication of an inference that the context effects are information-based. The information in an acoustic speech signal is about its gestural source in the vocal tract. That is, the structure in a speech signal is directly caused by the actions of the moving vocal tract; accordingly, to the extent that different actions of the vocal tract pattern the air pressure changes differently, structure in the acoustic signal provides information about its articulatory gestural source. It need not follow from this, of course, that listeners use acoustic structure in that way. However, there is reason to suppose that they do.

Across perceptual modalities, perceiving is the only means by which organisms can come to know the environment in which they participate as actors. But perception can be the means by which the environment is known only if stimulation at the sense organs—structured energy patterns in air and light, for example—serves not as something to be perceived and experienced in itself, but rather as information about the causal sources of its structure in the environment (e.g., Gibson, 1966, 1979). As visual perceivers, we *see* environmental sources of structure via reflected light; we do not see the structure in the light itself, even though it is the light and not the environment that stimulates the retina. We use the structure in reflected light to recover its environmental causes. Compatibly, as haptic perceivers we experience manipulable objects in the environment, not the skin and joint-angle deformations they cause. Accordingly, as auditory perceivers we should hear environmental sources of structure in acoustic signals, not the acoustic signals themselves, which should serve, instead, as information bearers. In speech, the sources of acoustic structure are linguistically significant

actions of the vocal tract (see Best, 1984, in press; Browman & Goldstein, 1986; Fowler, 1984, 1989; Fowler & Rosenblum, 1990; see also Liberman & Mattingly, 1985).

Setting aside for the moment the possible influence of perceptual learning, information in the acoustic signal about its origin in a sound-producing event in the environment—including vocal-tract actions—is available to any organism with an auditory system able to register the relevant acoustic structure. This includes prelinguistic infants, adult speakers from any language community and even nonhuman animals with appropriate auditory systems.

How, then, is perceptual elimination of coarticulatory influences of /l/ on following /g/ to be explained from this perspective? The /l/ in /alga/ is produced in part by creating a constriction between the tip of the tongue and the alveolar ridge of the palate. A /g/ is produced by creating a constriction between the back of the tongue and the soft palate. The forward constriction of the /l/ pulls the whole tongue forward, however. When production of the two phonetic segments overlaps, the constriction location for the following /g/ is fronted along the soft palate. The alveolar constriction, the soft-palate constriction and the causal effects of the former on the latter all have acoustic consequences. To the extent that the consequences are specific to those actions, the acoustic signal can specify those actions to a sensitive perceiver who then can ascribe the fronting to its source, the alveolar constriction. This information, if it is there at all, is as available to a prelinguistic infant as to a mature listener of any language community and even to a variety of nonhuman animals.[5]

As for the effect of learning a specific language on recovery of phonetic properties from an acoustic speech signal, our interpretation is similar to Mann's (1986). We have argued that listeners can recover information about vocal tract actions from acoustic speech signals. Mann refers to this as a "universal" level of perception to contrast it with a distinct, language-specific phonological level in which the linguistic significance of perceived gestures is appreciated. We will refer to the distinction in terms of attunement of attention, rather than perceptual levels. There is a mode of attending to acoustic speech signals that is available to listeners who participate in a particular language community and who have, therefore, discovered the linguistic significance, if any, of phonetic-gestural distinctions conveyed by an acoustic speech signal. This mode of attending is available to mature language users, but not to prelinguistic infants or to nonhuman animals (cf. footnote 5). While this linguistically-informed mode of attending to the signal is essential to linguistic interpretation of an utterance in the listener's native language (e.g., Best, Morrongiello, & Robson, 1981; Best, Studdert-Kennedy, Manuel & Rubin-Spitz, 1989), it may hinder explicit classification according to phonetic differences that are not phonologically distinctive in the native language (e.g., Werker & Logan, 1985). In making "l"-"r" classifications, Japanese listeners are impaired by their difficult-to-overcome tendency to ignore phonetic distinctions that are phonologically nondistinctive in their language. In contrast, all listeners can recover phonetic gestures of the vocal tract from the acoustic signal and can disentangle coarticulatory interactions among gestures, at least those that involve carryover, insofar as the acoustic signal specifies them. We suggest that prelinguistic infants eliminate coarticulatory influences of /l/ on /g/ precisely because the signal does specify the distinct articulatory correlates of /l/ and /g/ when the two segments are coarticulated.

Before concluding in this way, however, we consider an alternative, auditory, account of the findings of the present research that is also consistent with the inference that the context effects observed in this research are information-based. Mann considered this interpretation in her original article (1980), but not in her later one (1986), perhaps for a reason we outline shortly; two reviewers of the present manuscript requested that we consider the interpretation. We do so and explain why we consider it untenable.

The account ascribes the context effects of /l/ and /r/ on /d/-/g/ perception to auditory contrast. Constrast effects are widely observed in research obtaining perceptual judgments from subjects (see Warren, 1985 for a review), and on that basis alone, contrast might be considered a plausible or even likely cause of the present findings. In this instance, the high F3 of /al/ as compared to /ar/ may have a contrastive effect on judgments of the F3 transition of the following synthetic CV, leading listeners to judge it lower in frequency and hence more characteristic of /g/ than /d/. While the duplex perception experiment of Mann and Liberman (1983), cited earlier, rules out a locus for such an effect in the auditory system periphery, some contrast effects are thought to be more central in origin. In an example cited by one reviewer, Johnson (1944) found that immediately prior experience hefting weights gave rise to contrast effects on weight judgments; however, he observed informally that an interpolated weight

that subjects considered extraneous to the experimental setting—in particular, a book or chair that subjects might have moved during a rest break in the experimental proceedings—were "without apparent effect upon their scales of value based upon lifting the stimulus weights." (p. 436) If these informal observations are accurate and general, then perhaps the findings of Mann and Liberman, and hence of the present investigation can be explained in terms of contrast effects at a cognitive level. In particular, possibly in the research of Mann and Liberman (1983), the presence of context effects on the second syllable of the disyllables, but not on the isolated pitch glides, occurred because, as we suggested earlier, the pitch glides but not the disyllable's CVs were judged perceptually to constitute distinct objects from the influencing VCs.

An account in terms of auditory contrast makes qualitatively the same predictions concerning effects of spectral consequences of coarticulatory overlap on perception as does our proposed articulatory account. Acoustic effects of coarticulation are generally assimilatory, and contrastive effects of the coarticulating segment's acoustic consequences will always work to neutralize the perceptual effects of the assimilations. Qualitatively, this will also be the effect if listeners, as we suggest, ascribe coarticulatory influences to the coarticulating, rather than the influenced (target), phonetic segment.

Even so, for two reasons we discount the explanation of perception of coarticulatory context effects in terms of auditory contrast. The first reason concerns Mann's (1986) findings with Japanese listeners who were at chance in identifying /l/ and /r/, but who nonetheless exhibited context effects indistinguishable from those of English listeners and of Japanese listeners able to make the identifications. While findings of Mann and Liberman (1983) exclude a peripheral locus for any contrast effects just as they eliminate a peripheral locus for masking, findings by Mann (1986) with the first-mentioned group of Japanese listeners exclude a late, cognitive, locus—the locus at which Johnson's (1944) subjects would have excluded books and chairs from having a contrastive effect on weight judgments. Those listeners exhibited differential effects of context on phonetic segments that they could not label differentially. Accordingly, the contrast effects cannot arise early and they cannot arise late. There remains the possibility, of course, that contrast effects occur at some intermediate level of processing, less peripheral than the level at which duplex effects arise and more peripheral than that at which phonemic classifications occur. However, the articulatory account does not require such proliferation of processing levels, because it ascribes the effects to the relation between articulation and the acoustic signal, of which listeners are presumed to make use in perception. In articulation, phonetic segments are not discrete along the time axis; accordingly, listeners perceive a phonetic segment's domain to include its entire articulatory extent, insofar as it is specified acoustically and detectable auditorily.

A second reason to discount an explanation of the present findings in terms of auditory contrast is that the account does not explain the broader array of earlier findings concerning listeners' perception of coarticulated speech. It falls short in two domains, one relating still to spectral consequences of segment-to-segment coarticulatory overlap (classical coarticulatory effects) and the other to the acoustic consequences of other kinds of articulatory overlap.

In the literature, there are two complementary findings concerning listeners' perceptions as guided by spectral consequences of segment-to-segment coarticulatory overlap. One finding is exemplified by the present research. Listeners appear to eliminate effects of coarticulatory assimilations in their judgments of coarticulated segments, so that phonetic segments that are subject to coarticulatory overlap are both identified and discriminated as if the acoustic consequences of coarticulation were eliminated. Other research shows, however, that the acoustic effects of coarticulation are nonetheless perceptually effective as information for the coarticulating segment itself (e.g., Fowler, 1984; Fowler & Smith, 1986; Martin & Bunnell, 1981; Whalen, 1984). Indeed, in the research by Fowler (1984; Fowler & Smith, 1986), both findings are obtained using the same stimuli. That is, effects of coarticulatory assimilations appear to have been eliminated in discriminations of influenced segments, but nonetheless they serve as information for the coarticulating segment itself. Contrast effects can explain elimination of the effects of coarticulatory assimilations on perception of a target segment influenced by a coarticulating segment, but it is not obvious how they could put the effects back in elsewhere. Our account of perception, in fact, motivated the research cited above by Fowler, and predicted the obtained outcomes.

The second research domain in which the contrast account fails, in our view, has to do with listeners' perceptual handling of other kinds of ar-

ticulatory overlap, including prosodic and nonlinguistic variables that yield converging effects on fundamental frequency as reviewed in our introduction. The perceptual results are analogous to those in the literature just reviewed. That is, listeners judge intonation contours as if effects on the fundamental frequency contour of declination (Pierrehumbert, 1979; Silverman, 1987) and of segmental perturbations such as vowel height (Silverman, 1987) had been eliminated. Moreover, as in the literature on classic coarticulation effects, the "eliminated" effects are not eliminated in perception generally; they are eliminated only from listeners' judgments of the pitch melody of an utterance. Phonetic-segmental perturbations of the fundamental frequency contour of an utterance, including those due to variation in vowel height and consonant voicing, serve as information for their causes, namely vowel height (Reinholt-Peterson, 1986) and consonant voicing (Silverman, 1986), respectively. It is not obvious that a contrast account would handle even the elimination of the other-than intonational convergences on fundamental frequency from perception of the pitch melody, because articulatory overlap does not cause acoustic assimilation in these cases. Nor, analogous to the difficulties for the contrast account that we outlined relating to classic coarticulatory effects, does the contrast account appear to explain why the convergences, eliminated from one set of judgments (here, relating to intonational melody), do contribute to another set (that is, to judgments of vowel height or consonant voicing). An explanation that invokes recovery of the origins of the acoustic pattern in vocal-tract actions, however, does provide a unified account of the whole set of findings.

For these reasons, among others, we conclude that perception of coarticulated speech by adults and infants indexes their recovery of talkers' linguistically significant vocal-tract actions; it does not index auditory contrast.

REFERENCES

Bertoncini, J., Bijeljac-Babic, R., Jusczyk, P. W., Kennedy, L. J., & Mehler, J. (1988). An investigation of young infants' perceptual representations of speech sounds. *Journal of Experimental Psychology: General, 117*, 21-33.

Best, C. T. (1984). Discovering messages in the medium: Speech and the prelinguistic infant. In H. E. Fitzgerald, B. Lester, & M. Yogman (Eds.), *Advances in pediatric psychology* (Vol. 2, pp. 97-145). New York: Plenum.

Best, C. T. (in press). The emergence of language-specific phonemic influences in infant speech perception. In H. Nusbaum & J. Goodman (Eds.), *The transition from recognizing speech sounds to spoken words: Development of speech perception*. Cambridge, MA: MIT Press.

Best, C. T. McRoberts, G. W., & Sithole, N. N. (1988). The phonological basis of perceptual loss for non-native contrasts: Maintenance of discrimination among Zulu clicks by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 345-360.

Best, C. T., Morrongiello, B., & Robson, R. (1981). Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Perception & Psychophysics, 29*, 191-211.

Best, C. T., Studdert-Kennedy, M., Manuel, S., & Rubin-Spitz, J. (1989). Discovering phonetic coherence in auditory patterns. *Perception & Psychophysics, 45*, 237-250.

Browman, C., & Goldstein, L. (1986). Towards an articulatory phonology. *Phonology, 3*, 219-252.

Carden, G. C., Levitt, A., Jusczyk, P. W., & Walley, A. C. (1981). Evidence for phonetic processing of cues to place of articulation: Perceived manner affects perceived place. *Perception & Psychophysics, 29*, 26-36.

Crowder, R. G., & Repp, B. H. (1984). Single formant contrast in vowel identification. *Perception & Psychophysics, 35*, 372-378.

Diehl, R. L., & Kluender, K. (1989). On the objects of speech perception. *Ecological Psychology, 1*, 121-144.

Diehl, R., Elman, J. L., & McCusker, S. B. (1978). Contrast effects in stop consonant identification. *Journal of Experimental Psychology: Human Perception and Performance, 4*, 599-609.

Delgutte, B., & Kiang, N. Y. (1984). Speech coding in the auditory nerve: IV. Sounds with consonant-like dynamic characteristics. *Journal of the Acoustical Society of America, 75*, 897-907.

Eilers, R. E., & Oller, D. K. (1989). Conflicting and cooperating cues to final stop consonant voicing by infants and adults. *Journal of Speech and Hearing Research, 32*, 307-316.

Eimas, P. D. (1985). The equivalence of cues in the perception of speech by infants. *Infant Behavior and Development, 8*, 125-138.

Eimas, P. D., & Miller, J. L. (1980a). Organization in the perception of information for manner of articulation. *Infant Behavior and Development, 3*, 367-375.

Eimas, P. D., & Miller, J. L. (1980b). Contextual effects in infant speech perception. *Science, 209*, 1140-1141.

Elliot, L.L. (1971). Backward and forward masking. *Audiology, 10*, 65-76.

Fant, G., & Lindblom, B. (1961). Studies of minimal speech and sound units. *Speech Transmission Laboratory: Quarterly Progress Report, 2/1961*, 1-11.

Fitch, H., Halwes, T., Erickson, D. M.& Liberman, A. M. (1980). Perceptual equivalence of two acoustic cues for stop-consonant manner. *Perception & Psychophysics, 27*, 343-350.

Fowler, C. A. (1984). Production and perception of coarticulated speech in perception. *Perception & Psychophysics, 36*, 359-368.

Fowler, C. A., & Rosenblum, L. D. (1990). The perception of phonetic gestures. In I. G. Mattingly & M. Studdert-Kennedy (Eds.), *Modularity and the motor theory of speech perception*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Fowler, C. A., & Smith, M. R. (1986). Speech perception as "vector analysis": An approach to the problems of segmentation and invariance. In J. Perkell & D. Klatt (Eds.), *Invariance and variability of speech processes*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston, MA: Houghton-Mifflin.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston, MA: Houghton-Mifflin.

Grieser, D. A., & Kuhl, P. K. (1989). Categorization of speech by infants: Support for speech-sound prototypes. *Developmental Psychology, 25*, 577-588.

Harris, D. & Dallos, P. (1979). Forward masking of speech by the auditory nerve system. *Journal of Neurophysiology, 42*, 1083-1107.

Johnson, D. (1944). Generalization of a scale of values by the averaging of practice effects. *Journal of Experimental Psychology, 34*, 425-436.

Jusczyk, P. W., Pisoni, D. B., Reed, M., Fernald, A., & Myers, M. (1983). Infants' discrimination of a rapid spectrum change in nonspeech signals. *Science, 222*, 175-177.

Kent, R. D., Carney, P. J., & Severeid, L. R. (1974). Velar movement and timing: Evaluation of a model for binary control. *Journal of Speech and Hearing Research, 17*, 470-488.

Klatt, D. (1976). Linguistic uses of segment duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America, 59*, 1208-1221.

Klunder, K., Diehl, R., & Killeen, P. (1987). Japanese quail can learn phonetic categories. *Science, 237*, 1195-1197.

Krakow, R., Beddor, P., Goldstein, L. & Fowler, C. (1988). Coarticulatory influences on the perceived height of nasal vowels. *Journal of the Acoustical Society of America, 83*, 1146-1158.

Kuhl, P. K. (1979). Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories. *Journal of the Acoustical Society of America, 66*, 1668-1679.

Kuhl, P. K. (1980). Perceptual constancy for speech-sound categories in early infancy. In G. H. Yeni-Komshian, J. F. Kavanaugh, & C. A. Ferguson (Eds.), *Child phonology: Vol. 2, Perception* (pp. 41-66). New York: Academic Press.

Kuhl, P. K. (1983). Perception of auditory equivalence classes for speech in early infancy. *Infant Behavior and Development, 6*, 263-285.

Levitt, A., Jusczyk, P. W., Murray, J., & Carden, G. (1989). Context effects in two-month-old infants' perception of labiodental/interdental fricative contrasts. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 361-368.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21*, 1-36.

Mann, V. (1990, February 1). Personal communication.

Mann, V. A. (1980). Influence of preceding liquid on stop consonant perception. *Perception & Psychophysics, 28*, 407-412.

Mann, V. (1986). Distinguishing universal and language-dependent levels of speech perception: Evidence from Japanese listeners' perception of "l" and "r." *Cognition, 24*, 169-196.

Mann, V. A., & Liberman, A. M. (1983). Some differences between phonetic and auditory modes of perception. *Cognition, 14*, 211-235.

Martin, J. G., & Bunnell, H. T. (1981). Perception of anticipatory coarticulation effects in /stri,stru/ sequences. *Journal of the Acoustical Society of America, 69*, S92 (Abstract).

Miller, J. L., & Eimas, P. D. (1983). Studies on the categorization of speech by infants. *Cognition, 13*, 135-165.

Miller, J. L., & Liberman, A. M. (1979). Some effects of later-occurring information on the perception of stop consonant and semivowel. *Perception & Psychophysics, 25*, 457-465.

Moore, B. C.J. (1978). Psychophysical tuning curves measured in simultaneous and forward masking. *Journal of the Acoustical Society of America, 63*, 524-532.

Morse, P. A., Eilers, R. E., & Gavin, W. J. (1982). The perception of the sound of silence in early infancy. *Child Development, 53*, 189-195.

Oller, K. D. (1980). The emergence of the sounds of speech in infancy. In G. H. Yeni-Komshian, J. F. Kavanaugh, & C. A. Ferguson (Eds.), *Child phonology. Vol. 1: Production* (pp. 93-112). New York: Academic Press.

Pierrehumbert, J. (1979). The perception of fundamental frequency declination. *Journal of the Acoustical Society of America, 66*, 363-369.

Reinholt-Peterson, N. (1986). Perceptual compensation for segmentally-conditioned fundamental-frequency perturbations. *Phonetica, 43*, 31-42.

Repp, B. H. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin, 92*, 81-110.

Silverman, K. (1986). F0 segmental cues depend on intonation: The case of the rise after voiced stops. *Phonetica, 43*, 76-91.

Silverman, K. (1987). *The structure and processing of fundamental frequency contours*. Unpublished doctoral dissertation, Cambridge University, Cambridge, UK.

Smith, R. L. (1977). Short-term adaptation in single auditory-nerve fibers: Some post-stimulatory effects. *Journal of Neurophysiology, 40*, 1098-1112.

Stark, R. E. (1980). Stages of speech development in the first year of life. In G. H. Yeni-Komshian, J. F. Kavanaugh, & C. A. Ferguson (Eds.), *Child phonology. Vol. 1: Production* (pp. 73-92). New York: Academic Press.

Summerfield, A. Q., & Haggard, M. P. (1977). On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants. *Journal of the Acoustical Society of America, 62*, 435-448.

Warren, R. M. (1985). Criterion shift rule and perceptual homeostasis. *Psychological Review, 92*, 574-584.

Werker, J., & Logan, J. (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics, 37*, 35-44.

Whalen, D. H. (1983). Vowel information in postvocalic fricative noises. *Language and Speech, 26*, 91-100.

Whalen, D. H. (1984). Subcategorical mismatches slow phonetic judgments. *Perception and Psychophysics, 35*, 49-64.

FOOTNOTES

Perception & Psychophysics, 48, 559-570 (1990).

†Dartmouth College, Hanover, New Hampshire.

††Wesleyan University, Middletown, Connecticut.

†††Department of Psychology, Stanford University, Stanford California.

[1]The latter authors reported a case of failure of perceptual equivalence, in both infants and adults, for the contributions of release burst and vowel length to perceived voicing of a final stop. However, these two acoustic cues do not result from a unitary phonetic gesture, and so would not be expected to show perceptual equivalence.

[2]We thank Virginia Mann for loaning us her stimuli.

[3]The authors also completed the test, but their data were not included in the final analyses.

[4]We do not know why such an asymmetry should occur. However, perhaps if /l/ pulls /g/ forward, /g/ does not correspondingly pull /l/ back very far due to /l/'s fixed, and /g/'s sliding, place of articulation along the palate.

[5]In this respect we disagree with Kluender, Diehl, and Killeen (1987) who conclude that the Japanese quail's ability to categorize novel CV syllables based on the initial consonant is not attributable to perceived articulation. ("On what basis do these quail correctly categorize new tokens? The possibility that their categorizations are based on a knowledge of articulatory commonalities can be excluded."—p. 1196) We would not be surprised to find that quail could categorize novel instances of active humans into the classes "walking" or bipedal "hopping;" moreover, if they could, we would presume that the categorizations were based on the perceived distal events of people either walking or hopping as those events are conveyed by information in reflected light to the eye. It seems to us no less plausible to suppose that quail can categorize novel instances of utterances into classes /d/-initial and /b/-initial based on the perceived distal events of vocal tract-like systems producing those consonants as those articulations are conveyed by information in acoustic speech signals.

Haskins Laboratories Status Report on Speech Research
1991, SR-105/106, 107-140

Extracting Dynamic Parameters from Speech Movement Data

Caroline L. Smith,† Catherine P. Browman, Richard S. McGowan, and Bruce Kay††

A quantitative characterization of movement trajectories, using the parameters of a linear second-order dynamical system, was developed in order to make possible comparisons among classes of movements, in particular classes defined by linguistic factors such as syllable position, stress and vowel quality. Testing on simulated data with known parameter values showed that natural frequency, the parameter of principal interest, could be estimated with a mean error of no more than 5%, if only frequency and d.c. level were allowed to vary during the fitting procedure. The movement trajectories were divided into sections ("windows") in two ways: at successive displacement peaks and valleys, and at the right edge of plateau regions around such extreme values. The damping ratio was estimated by fitting the data at damping ratios that were varied from 0.0 to 1.0, then selecting the fit of each window of data that had the least error. Measured articulatory data were also analyzed using fits made with a single fixed damping ratio, rather than the variable damping ratios obtained by using the least error criterion. Although the values obtained for natural frequency increased, on average, as the damping ratio increased, the patterns of the effects of the factors remained stable at all the tested damping ratios.

1 INTRODUCTION

When studying articulatory movement data, the problem of measuring a continuously varying signal must be faced. Analysis of such a signal requires a quantitative description of the spatial and temporal properties of the movement. A useful description of the movement would have fewer degrees of freedom than the data and capture its characteristics as a member of a class of movements, while accurately representing its idiosyncratic properties. Such a description should also facilitate comparison among classes of related movements. A kinematic description of a movement, which connects displacements to specific times, may fail to capture the underlying relatedness among class members; for these reasons, a more abstract, dynamical description seems preferable. Various investigators have modelled the movements of the speech articulators as a dynamical system (e.g., Sonoda & Kiritani, 1976; Fowler, Rubin, Remez, & Turvey, 1980; Ostry, Keller, & Parush, 1983; Ostry & Munhall, 1985; Browman & Goldstein, 1985; Kelso, Saltzman, & Tuller, 1986; Saltzman & Munhall, 1989).

The equations of motion for a dynamical system represent changes in spatial variables over time by stating a relationship among variables of motion that remains constant over time. An example of a simple dynamical system is the mass-spring system, modelled by a linear, second-order dynamical system. This type of dynamical system has been shown to produce trajectories with the connection between displacement and peak velocity that is characteristic of (reiterant) speech movements over changes in stress and speaking rate (Tuller, Harris, & Kelso, 1982; Ostry & Munhall, 1985; Kelso, Vatikiotis-Bateson, Saltzman, & Kay, 1985; Vatikiotis-Bateson, 1988). In particular, changes in stress have been treated as changes in articulator stiffness (Munhall, Ostry, & Parush, 1985; Kelso et al., 1985;

The authors would like to thank Greg Burton and Craig Cooper, who performed much of the preliminary testing and data analysis for this project, and Katherine Harris and Ignatius Mattingly, who gave helpful comments on an earlier version of the manuscript. This work was supported by NSF grant BNS 8820099 and NIH grants HD-01994 and NS-13617 to Haskins Laboratories.

Browman & Goldstein, 1985), since change in the displacement/peak velocity relationship can be modelled in a second-order system as a change in articulator stiffness (Cooke, 1980).

Modelling articulatory data as a dynamical system relates it to a well-defined system with a constrained description. Thus a relatively small number of parameters is needed to specify a particular trajectory within a given system. An overall comparison of two classes of movements can be made by comparing the two sets of parameters that represent the classes rather than comparing two sets of trajectories directly. The work reported here investigates how the parameters for a second-order dynamical system, particularly stiffness and damping, reflect changes in linguistic factors such as stress, syllable position, and vowel quality. The first part of this paper will discuss the procedures used to verify the accuracy with which trajectories can be modelled by a particular second-order system. In the second part of the paper, we will show that the extracted parameters capture systematic effects of the linguistic factors on articulatory movements.

In order to extract the dynamic parameters that characterize the articulatory trajectories, we have developed a computer program (PARFIT) to identify these parameters using nonlinear least-squares curve fitting. The system used as a model is the mass-spring, whose equation is

$$m\ddot{x} + b\dot{x} + k(x - x_0) = 0 \qquad (1)$$

where m=mass, b=damping, k=stiffness, and x_0 =rest position. (In the work reported here, mass is assumed to be equal to 1.) The parameters extracted from the movement data correspond to the coefficients of the trigonometric form of the solution, shown in (2) below, which can be related analytically to the mass-spring equation above.

$$x(t) = e^{\alpha t}(A\cos\beta t + B\sin\beta t) + d.c. \qquad (2a)$$

$$= \sqrt{A^2 + B^2}\; e^{\alpha t}\cos(\beta t - \theta) + d.c. \qquad (2b)$$

where α=growth, β=observed frequency, d.c. is the dc offset or constant level, A and B are a function of two selected values from the data (to be discussed later), and θ is determined by A and B. The parameters in equations (1) and (2) are related as follows

$$\alpha = \frac{-b}{2},\ \beta = \frac{\sqrt{4k - b^2}}{2},\ \text{and d.c.} = x_0 \qquad (3)$$

when mass = 1. Observed frequency (β) is related to natural frequency (ω_0) and damping ratio, itself a function of damping and natural frequency:

$$\omega_0 = \sqrt{\frac{k}{m}},\ \beta = \omega_0\sqrt{1 - (\text{d.rat.})^2},\ \text{and d.rat.} = \frac{b}{2\omega_0} \qquad (4)$$

A particular set of dynamical parameters describes a particular dynamical system that fits the articulatory data for a period of time. But the system state controlling the articulatory movements does not remain the same indefinitely: continuous movement is associated with a set of phonetically discrete units, or "gestures" (see Browman & Goldstein, 1989). The system parameters change between different phonetic units. Thus the movements themselves must be divided up into the sections or "windows" that correspond to different control regimes for discrete phonetic units (Browman & Goldstein, 1985). How to section the continuous movements, and how the sectioning may relate to phonetically significant units, is an empirical issue in itself, with important effects on the extracted parameter values. The effects of different ways of dividing data into segments, or "windows," is discussed below.

At Haskins, a computational model of linguistic gestures is being developed that uses a second-order, damped mass-spring system to generate model articulator trajectories (Browman, Goldstein, Saltzman, & Smith, 1986; Saltzman, Rubin, Goldstein, & Browman, 1988; Browman & Goldstein, 1990). Since the data analysis reported here was intended to uncover characteristic patterns for use with this model, the requirements of the model (e.g. the use of a second-order system) were a constraint on the type of dynamical system chosen for the parameter extraction. The model also oriented the data analysis; for example, since one simplifying assumption currently implemented in the model is that the damping ratio is constant, we investigated how well a constant damping ratio could characterize the articulatory movements.

2 ANALYSIS OF SIMULATED DATA

To interpret the characterization of the data provided by the extracted parameters, we need to know how reliable the results are, not only in the sense of how close are the values provided by the program to the "true" parameters of the data, in the case that they are independently known, but also how well the parameters mirror changes in the data. That is, how consistently do the parameter values characterize the data? In order to test the reliability and consistency of the program, we first tested it on simulated data.

PARFIT returns the squared error of the fit of each window. However, that information alone does not show how a given error size translates

into differences between fit and data curves. It is also necessary to know how accurately the size of the squared error can be used to select the set of parameter values that best fits the data curve. To answer this question we need a way to compare the parameters supplied by PARFIT with the values that actually gave rise to the data. This comparison is possible when simulated data created with known parameter values are analyzed.

2.1 Simulated data

Four sets of simulated data were created that were intended to capture the range of values that were expected, on the basis of previous experience, in the measured data. Two of these sets had a natural frequency of 4 Hz and two of 12 Hz. Of each of these pairs, one set had a displacement amplitude in the first window of approximately +/-150 and the other +/-1000 machine units. For each combination of frequency and displacement amplitude, eleven data curves were created at damping ratios from 0.0 to 1.0 in increments of 0.1. Examples of the shapes of the curves for four of these damping ratios in the 4 Hz, low amplitude set are shown in Figure 1. There were a total of 44 data files, each with a unique combination of frequency, displacement amplitude and damping ratio.

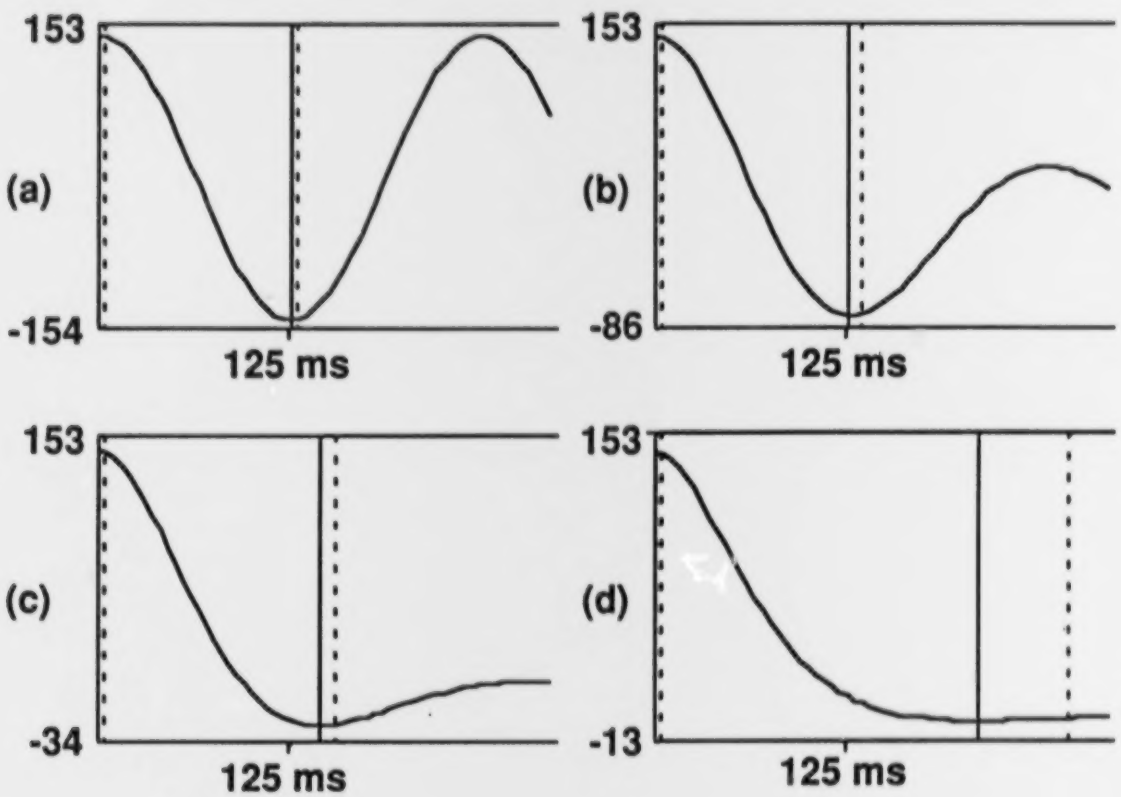


Figure 1. **Simulations at four damping ratios, (a) undamped, (b) .2, (c) .5, and (d) .8, with Peak (solid lines) and CV (dotted lines) windows marked. The values on the vertical axis give the range of the display; the maximum (or minimum) in each figure is 10 machine units larger (or smaller) than the extreme value of the curve. The total duration displayed in each figure is 300 ms.**

To compare possible effects of different methods of sectioning curves, the trajectories were divided into windows in two different ways (see Figure 1). One was the customary peak-to-valley size unit, referred to here as Peak windows, where each window (portion of the curve) extends from the midpoint of a peak (or valley) to the midpoint of the next valley (or peak). An alternative to this method, referred to as "CV windows", was also tested. This method was intended to include the relatively flat plateau regions around displacement extrema with the regions of movement between these plateaus. One possible way of accomplishing this result is to divide the trajectories at points of peak velocity (Browman & Goldstein, 1985); however, such divisions do not correspond in any obvious way to linguistically significant aspects of the articulation. Another possible way is to divide the trajectories into plateau regions around position peaks or valleys, with relatively straight paths between these regions. However, preliminary investigations revealed that analyzing the straight portions was problematic, since such regions could be fit by curves with a very wide range of parameter values. Therefore, the method that was ultimately used included the region around a peak (or valley) as well as the movement towards it (see Figure 1). Each CV window extended from the right edge of a plateau region around a peak (or valley) to the right edge of the next valley (or peak). In the simulations, the plateaus begin or end within 1% of the range of amplitude after the extreme peak or valley of the trajectory. The parameters obtained from two sets of analyses, one using these CV windows and the other the Peak windows, were then compared.

The analyses below treat only the results of the analysis of the first window in each simulation, because each window in measured articulatory data is assumed to be the first window of a new regime; thus, the first window of the simulated data should be most similar to the measured articulatory data. The first window in the simulations was always an "opening" (lowering) movement.

2.2 Procedure

PARFIT fits curves using a multidimensional Newton's method with a least-squared error criterion. That is, the program attempts to find a set of parameters such that the sum of squares of the differences between the positions generated by those parameters and the corresponding position data points is minimized. Further details about the algorithm used for fitting and tests of its validity can be found in McGowan, Smith, Browman, and Kay (1990). When all five parameters of equation (2a) that characterize a fit (frequency, constant level, growth, and the amplitude coefficients of the cosine and sine) were allowed to vary, the results obtained were unstable; that is, the parameters tended to trade off against one another, so that a small change in the curve resulted in large changes in two or more parameters. Therefore, the number of parameters to be fit at the same time was reduced to three (frequency, constant level, and growth) by treating the coefficients of the cosine and the sine in the equation for the curve, equation (2a), as functions of the above three parameters plus a pair of values selected directly from the data. Two alternative pairs of derived data values were tested. In one, referred to throughout as the "Initial condition", or Initial fit, the displacement and velocity of the data curve at the first point in the window being analyzed were selected. In the other, referred to throughout as the "Boundary condition", or Boundary fit, the displacements at both the initial and final points of that window were selected.

Tests were run using the above three parameters (frequency, constant level, and growth), fitting the simulated data. These tests (McGowan, Smith, Browman, and Kay 1988) revealed that although the fits might be fairly close to the data (a squared error of approximately $1 \times 10E\text{-}7$), the variance of the parameter values could be of the same magnitude as the values themselves, except that the variance was a little smaller for frequency in Boundary fits. In addition, the remaining parameters still had the potential to trade off against one another (indicated by large covariance among them). These results suggested that even three parameters may be too many to get stable results. To cut back to only two parameters, analyses were run holding damping ratio constant, and allowing only frequency and constant level to vary. In this procedure, the damping ratio of the data was estimated by analyzing the data at many different damping ratios and selecting the one providing the fit with the lowest error.

Each of the 44 simulated data files was sectioned using both CV and Peak windows. Then, each file was analyzed with PARFIT for each Window Type using both the Initial and the Boundary conditions with the damping ratio held fixed at 11 values in .1 increments from 0.0 (undamped) to 1.0 (critically damped). This meant that each of the 44 simulated data files was analyzed, or fit, 44 times. The one exception was

files labeled with CV windows, when analyzed using the Boundary condition. These files were not fit with critical damping because of the inherent contradiction between the curve after the displacement extremum in CV windows, and the nature of critical damping. A critically damped curve will not turn away from its displacement extremum unless it has high initial velocity, and the initial velocity in all windows was very low because of the criteria used to section the curves into windows. This contradiction was only a problem with the Boundary fits because in that condition both endpoints of the window must be fit exactly, whereas in the Initial condition the fit curve may diverge from the final value in the window.

To control how closely the fit must correspond to the (simulated) data, the lowest desired error between fit and data was specified. (The error criterion was increased when no fit could be found using a more restrictive lower error.) Since a change in the size of the error criterion can result in different fits being found, the smallest error criterion used in fitting the simulated data was chosen on the basis of preliminary analyses of measured articulatory data, although because the simulated data were perfectly smooth and noise-free, many fits had errors much smaller than the criterion.

One of the general concerns in finding fits was that the program sometimes found parameters that, although they fit the data with relatively low error, were clearly inappropriate. For example, sometimes the value for the parameter constant level, which represents the amount of d.c. offset of the target, was far greater than any amplitude value reached by the data curve. This kind of fit occurred if the program analyzed the data curve as corresponding to only a small portion of the underlying cycle of the model curve. To exclude such fits, constraints were placed on the ranges of possible values for the parameters. The constraints permitted parameter values to vary over a larger range of values than was expected to occur in the data. For example, the observed frequency was constrained to be between 0 and 20 Hz; the range of possible values for growth was limited to between -200 and 5 sec^{-1}, allowing only damped curves, with an allowance for a small positive amount of growth; and the constant level could not exceed 1.1 times the amplitude range of the window. (For opening (lowering) gestures, this value was the minimum allowable d.c. value; for closing (raising) gestures, the maximum. If the multiplication factor were exactly 1.0, then in the case of critically damped data, this would imply that the d.c. value—the rest position or target—was expected to be exactly the extreme amplitude value in the window. 1.1 allows a little leeway in this assumption, i.e., the target can be a little beyond the observed extrema. This assumption is most restrictive in the critically damped case.)

In addition, we could check more directly whether the fit curve was a reasonable portion of the underlying curve. The program provides values for the phase angle, which is the number of degrees of the curve that were fit, defined in terms of its natural frequency. Windows being fit were typically approximately half-cycles of the observed frequency. As the damping (ratio) increases, for example, for peak-to-valley windows, the observed movement extremum—180 degrees of the observed frequency—corresponds to an increasing phase angle of the underlying natural frequency. In addition, as the estimated frequency increases, the estimated phase angle also increases, for a given movement curve. Thus, expected values of the phase angle can be calculated as a function of damping ratio and natural frequency. In our testing process, for simulations with damping ratio less than .9 the correct regime always had the phase angle within 10% of the correct value. In this way the phase angle provided us with a check on the accuracy of the criterion of least-squared error for selecting the best fit, since a highly divergent phase angle suggests the fit is not close to the data, particularly if the error for the fit is also relatively high.

2.3 Selecting among the fits

With the goal of finding fits that gave accurate estimates of the parameters, the criterion chosen to select among fits was the size of the squared error. Among fits of the same data file made under similar conditions, the hypothesis was that the one with the smallest squared error (referred to as "least error") should be the most accurate fit. This hypothesis was tested by seeing if smaller error reliably correlated with a fit that provided parameter values closer to the correct values for the simulation than fits with larger error. Fits whose error sizes were close to each other should not differ greatly. These results will be discussed below.

Selected fits were also compared qualitatively using graphical representations to see how they diverged from the simulated curve. An example of such a comparison is shown in Figure 2, which illustrates the results of using increasing damping ratio to fit an undamped curve using the Boundary condition. At damping ratios .2 and .5, the fits shown in Figure 2(a) and 2(b) diverge

slightly from the simulated data curve, but they are quite close and in both cases fit the simulation, which consisted of half a cycle of an undamped sinusoidal curve, with approximately half of one cycle of the fit curve. As is apparent from Figure 2(c), the close fit breaks down at fit damping ratio .8. In this case, the fit curve diverges drastically from the simulation, fitting the endpoints exactly (because the fit was run using the Boundary condition), but using much more than half of one cycle of the fit curve to fit the half-cycle simulated data window. Since the fit must meet the simulation at the endpoints, an alternative strategy for the fitting procedure would be to fit the undamped half-cycle simulation with much less than half a cycle of a damped curve. Both of these possible strategies occur in fitting when the damping ratios of the data and the fit are incompatible. Inaccuracy of the type shown by this fit is eliminated by the least error criterion, which would select for this curve a fit other than the one generated by a damping ratio of .8. In fact, it will be shown that the least error criterion selected a fit generated by a damping ratio no more than .1 away from the actual damping ratio of any simulation.

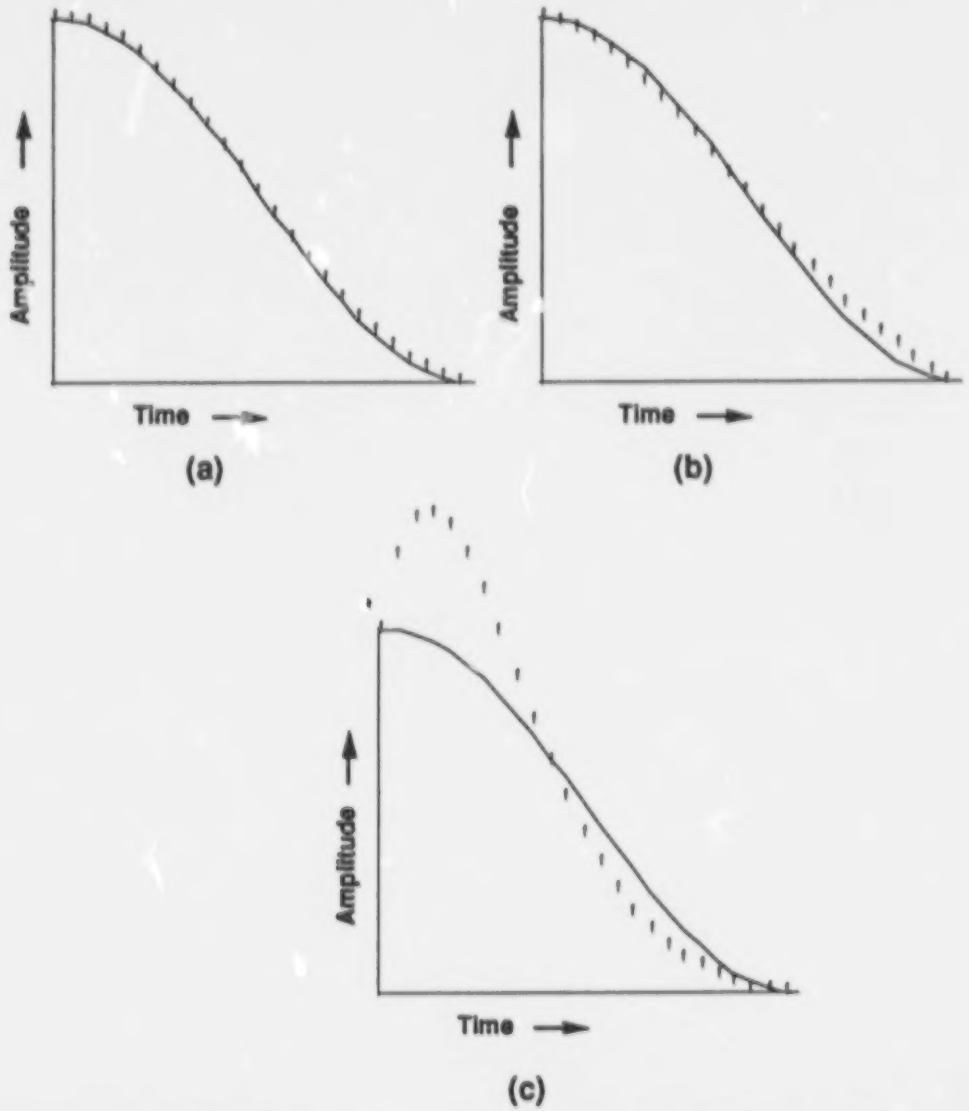


Figure 2. The solid line represents a half-cycle of an undamped, simulated data curve (damping ratio 0.0). The f's show fits of this curve using the Boundary condition (a) with a damping ratio of .2, (b) with a damping ratio of .5, and (c) with a damping ratio of .8.

In order to determine how accurately the damping ratio of a curve could be estimated, for each simulated data file and combination of window type and Initial/Boundary condition the damping ratio of the fit that had the smallest error was selected. Table 1 shows how many times (out of 11 data files per condition) the fixed damping ratio that gave the smallest error was in fact the actual damping ratio of the simulated data.

Table 1. *Number of fits where the damping ratio with the least error was the actual damping ratio of the simulated data. (total possible = 11).*

simulated data		type of fit condition and window			
Frequency	Amplitude	Boundary		Initial	
		CV	Peak	CV	Peak
4 Hz	low	7	4	4	11
4 Hz	high	3	5	3	11
12 Hz	low	8	6	4	11
12 Hz	high	9	8	2	11

The Peak Initial fit—that is, fits using the Initial condition and Peak windows—are artificially accurate, because in these simulations, the first window started at the first data point, making it impossible to calculate the velocity using a centered difference as was normally done. Therefore the exact value for the velocity (=0) was provided, making these fits more accurate than they would normally be. The accuracy of Peak Initial fits cannot, as a result, be compared directly to the accuracy of the other types of fit. (See McGowan, Smith, Browman, and Kay (1990) for more details of these analyses.)

Figure 3 shows the distribution of least error fits around the correct value for the damping ratio. The simulations were grouped into those with "low" damping ratio, 0.0 - 0.8, and "high" damping ratio, 0.9 and 1.0. The rationale for this grouping will be discussed below. Peak Initial fits were excluded from this comparison because they always gave the correct answer; therefore the graph includes only 44 CV Initial fits and 88 Boundary fits. Although the least error criterion did not always select the correct regime, it was usually close: the graph shows that 88% of the least error fits were within +/- .1 of the correct damping ratio.

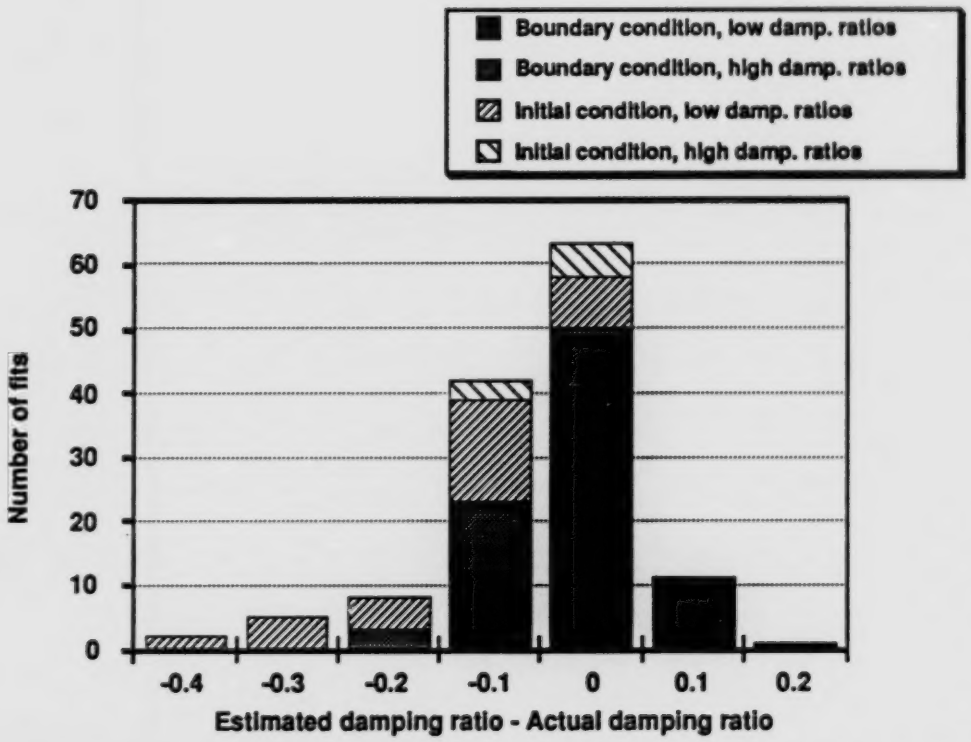


Figure 3. **Out of 132 (44 CV Initial and 88 Boundary) fits, the number of fits for the indicated difference between the actual damping ratio of the simulation and the estimated damping ratio (using least error). When the least error fit gave too low an estimate of damping ratio, the value is plotted as negative. When the least error fit gave too high an estimate of damping ratio, the value is plotted as positive.**

The least error fits in which the damping ratios were more than +/- .1 away from the correct damping ratio tended to fall into two groups. Least error fits with damping ratio more than .2 away from the correct damping ratio occurred only in Initial CV fits; moreover, these instances all involved simulations at damping ratios of .6 or less. Conversely, the Boundary condition came closer to the correct damping ratio for data with lower damping ratios and selected less accurately for data at .9 and 1.0. This contrast in the behavior of the two conditions suggested that perhaps the different conditions should be used for different damping ratios of the simulated data. When the Boundary condition was used for fits whose damping ratio was 0.0 to 0.8, and the Initial condition for fits with damping ratios of 0.9 and 1.0, 87 of the 88 data files had the least error fit within +/- 0.1 of the correct damping ratio. These results suggest that the Boundary condition should be specified for analyses using fit damping ratios between 0.0 and 0.8, while the Initial condition should be specified for analyses using fit damping ratios of 0.9 or 1.0.

Further evidence suggesting that Boundary fits are preferable to Initial fits at lower damping ratios was found in tests performed on simulated data with damping ratios of .4 and .8 to see how sensitive the fits are to noise (McGowan, Smith, Browman, and Kay (1990)). In simulated data with added noise, the damping ratio of the least error fit using the Boundary condition was more consistently close to the damping ratio of the data than the least error fit using the Initial condition. Also, the parameters obtained using the Boundary condition had covariance matrices whose elements were smaller relative to the parameter values themselves than parameters obtained using the Initial condition. This test constituted further evidence that the Boundary condition was more stable in fits at damping ratios up to .8.

2.4 Accuracy of values for natural frequency

With a criterion for choosing a fit provisionally established, we continued to test it by investigating the accuracy of the natural frequency estimates for all fits whose damping ratio fell within +/- 0.1 of the correct damping ratio, since that was the limit of the accuracy with which the damping ratio of the fit could estimate the damping ratio of the simulated data. Comparison between Initial and Boundary conditions with regard to accuracy in estimating natural frequency showed that there was an interaction between condition and damping ratio of the fit similar to what was found in estimating damping ratio. When the optimal combinations of these interactions were used, on the average the estimated natural frequencies were within 5% of the correct value.

An examination of trends in the natural frequency estimations showed that the accuracy of its values was more sensitive to the accuracy of the estimated damping ratio as the damping ratio increased. This means that an error in damping ratio will have more effect on the estimated value of frequency for highly damped data than for relatively underdamped data. However, the tests reported here suggest that any error in estimating damping ratio should be no more than .1, even for highly damped data (assuming the optimal Initial/Boundary conditions are used).

2.4.1 Initial versus boundary conditions

To compare the accuracy of the values for natural frequency provided by the least error fits, the mean percentage error of all the fits whose damping ratios were within +/- .1 of the correct damping ratio is plotted in Figure 4 for each damping ratio of the simulations. (Recall that the Initial condition Peak window fits, which have very low error, were artificially accurate.) At lower damping ratios, the error of the Boundary fits is somewhat smaller than the error of the Initial fits for CV windows. This difference becomes small in simulations with damping ratios of .6 and .7, and reverses at damping ratios of .8 and above. Indeed, at damping ratios of .9 and 1.0, the error of the Boundary fits becomes dramatically higher, while the error of the Initial fits remains at approximately the same level. Clearly Initial fits are to be preferred at damping ratios of .9 and 1.0, with a mean error of 6.1%.

For simulations with damping ratios of 0.9-1.0, the accuracy for Initial fits was clearly much better than that for Boundary fits. However, for less damped simulations, Boundary and Initial conditions gave more similar results. Since Initial fits with Peak windows were artificially accurate (as discussed in section 2.3), a more realistic comparison of Initial and Boundary fits for the less damped data would eliminate the Peak windows, and compare only fits made using CV windows, as shown in the first row of Table 2. Comparing only these fits, an analysis of variance (using BMDP 4V) showed that the mean error for Boundary fits was significantly lower than for Initial fits ($F=15.14$, $p<.001$), supporting the preference for Boundary fits for data fit at damping ratios of 0.0-0.8.

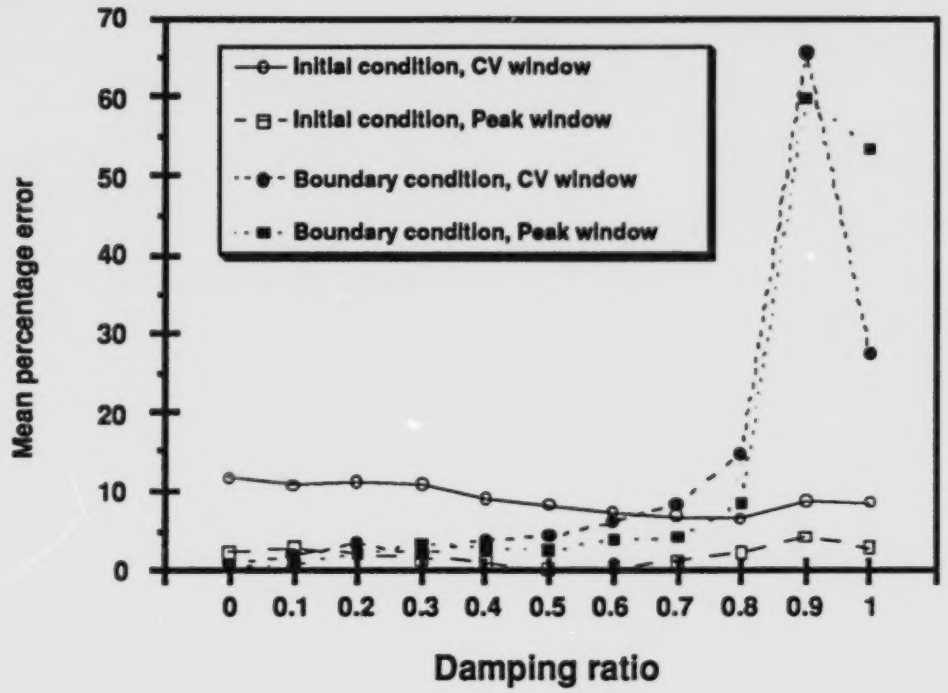


Figure 4. **For simulations at each damping ratio using all fits whose damping ratios were within +/- .1 of the simulation, the absolute value of the mean percentage error in the estimates of natural frequency provided by the fits.**

Table 2. *CV windows only: Absolute value % error in natural frequency. Fits at damping ratios within +/- 0.1 of correct damping ratio.*

	Boundary		Initial	
Simulation	Mean	S.D.	Mean	S.D.
0.0-0.8	5.2	8.7	9.1	5.2
0.8 only	8.45*	4.84	6.49	5.14

*excluding outlier value of 82.5%

The results discussed above showed a distinction in accuracy of estimates of natural frequency between Initial and Boundary fits for damping ratios below .8 and those at damping ratios of .9 and 1.0. In section 2.3 we discussed how the higher accuracy and lower variability of estimations of damping ratio in the Boundary condition fits suggested that it might be preferred over the Initial condition at the lower damping ratios, up through .8. To test whether .8 should indeed be included with the lower damping ratios, using the Boundary condition, further comparisons were made between Initial and Boundary condition fits in simulations whose damping ratio was exactly .8. Again, the accuracy of the estimates of natural frequency were compared only in fits whose damping ratios were within +/- 0.1 of the correct damping ratio, that is 0.7-0.9. Since the Initial Peak fits were (misleadingly) artificially accurate, a comparison between Boundary and Initial was made using CV windows only. For the .8 simulations, in CV windows, there was not a significant difference in the accuracy of frequency values between Initial and Boundary fits ($F=0.88$, $p=.36$, excluding one outlying Boundary value), as shown in the second row of Table 2.

Therefore, based on the damping ratio analyses we included fit damping ratios of .8 in the lower group, and in all future analyses used the Boundary condition for fits whose damping ratios were .8 or lower, and the Initial condition for fits with damping ratios of .9 or 1.0. By taking advantage of this interaction between the condition options and fit damping ratios, we avoided using types of fits with intolerably high variability. Taking the fit with the least error from the combination of Boundary fits at low damping ratios and Initial fits at high damping ratios also promised to be a reasonably accurate method for estimating the actual damping ratio and natural frequency of the data being fit.

2.4.2 Peak versus CV windows

Having decided how to use the condition options, we then compared the two choices of window type, CV and Peak windows, to see whether either one gave more accurate values for natural frequency for the simulated data. The fits used in this comparison were only those with the Boundary condition at damping ratios up through .8, and the Initial condition at .9 and 1.0, and only those whose damping ratio was within +/- .1 of the simulation being fit. The percentage error in natural frequency values obtained using the two window types is shown in Figure 5. Using these fits in an analysis of variance (with BMDP 4V), Peak windows gave significantly more accurate values ($F=9.18$, $p<.01$). However, again these results were skewed by the unrealistic accuracy of the Peak Initial fits. To avoid this problem, another analysis of variance was run using only the Boundary fits from the above data (i.e. within +/- .1 of the correct damping ratio for data at damping ratios 0 - 0.8). In this analysis, fits using Peak windows still gave frequency values that were closer to the correct value, but the difference was only marginally significant ($F=4.44$, $p<.04$). Since both window types provided estimates of damping ratio to within .1 of the correct value of the simulations, and mean estimates of natural frequency within 10% of the correct value (when used in conjunction with the appropriate choice of Boundary or Initial condition depending on the damping ratio of the attempted fit), both were used in the analyses of the articulatory data described below, with the goal of investigating the kinds of differences in the results obtained in the two cases.

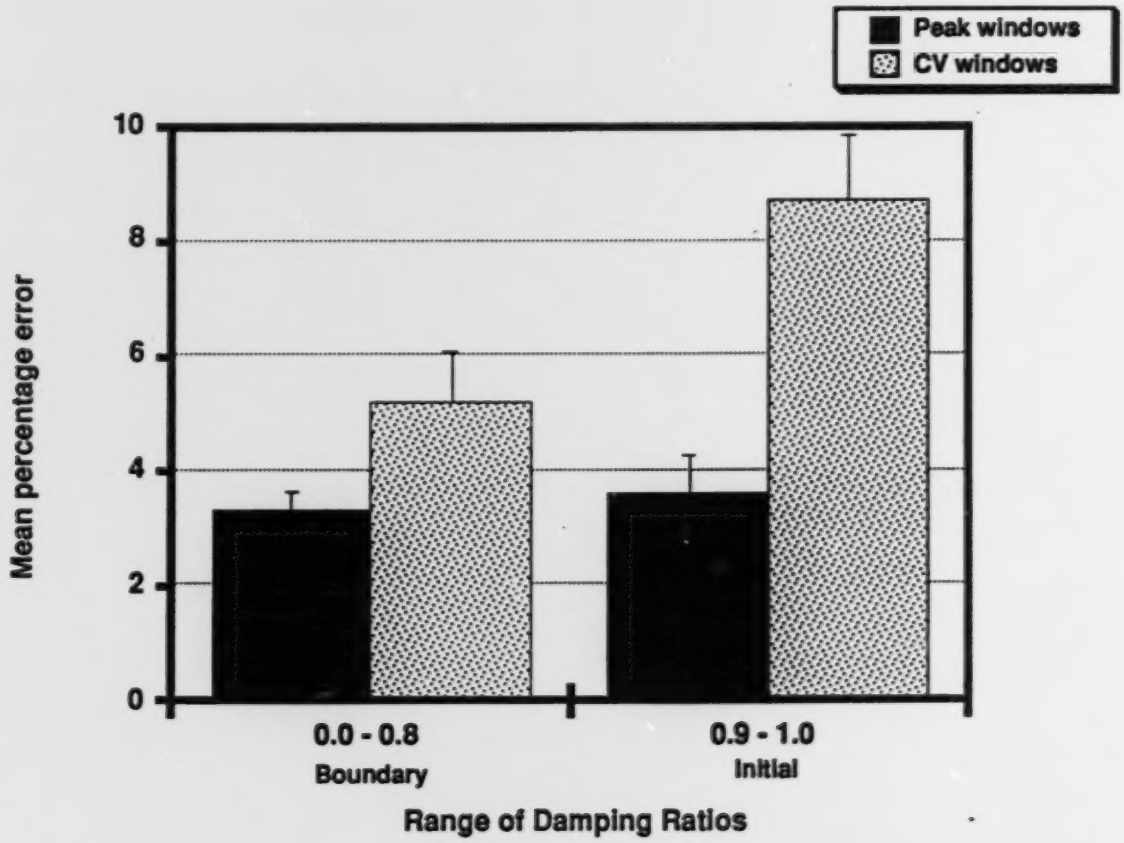


Figure 5. Comparison between Peak and CV windows of the absolute values of the mean percentage error in natural frequency for fits of simulations having damping ratios up through .8 (with the Boundary condition) or .9 and 1.0 (with the Initial condition). Only fits whose damping ratios were within +/- .1 of the simulation were used. The magnitude of one standard error is shown by the error bar on each column.

3 ANALYSIS OF ACTUAL ARTICULATORY DATA

Working with the simulated data enabled us to develop a technique for analyzing actual articulatory data with PARFIT. To summarize, a data file was fit with Boundary condition fits at 9 different damping ratios (0.0 -0.8) and with Initial condition fits at 2 damping ratios (0.9 and 1.0). Both window types (Peak and CV) were used, making a total of 22 fits per data file. The fit with the least error for each window type was selected. This analysis technique could then be used on measured articulatory data whose parameter values are, of course, unknown.

We wanted to know to what extent the parameters extracted by PARFIT provide a way to quantify the effects that linguistic factors such as stress, syllable position and vowel quality have on articulatory gestures, and whether these effects are revealed in the actual values of the parameters, or just in patterns that remain stable across, for example, different damping ratios. For instance, it is well known that unstressed gestures tend to be of shorter duration and show smaller movement amplitudes than stressed ones (e.g., Kent & Netsell, 1971; Kozhevnikov & Chistovich, 1965; Ostry, Keller, & Parush, 1983; Tuller, Harris, & Kelso, 1982; Zawidzka, 1981). PARFIT should reflect this tendency by assigning unstressed gestures higher frequency values, regardless of damping ratio. Primarily we were interested in how the linguistic factors and the type of window affected the values of natural frequency obtained from the least error fits. From the tests of the simulated data, we expected these results to be reasonably accurate.

3.1 Data collection

In the utterances whose movements would be analyzed by PARFIT, a small set of linguistic factors was varied, with two values for each factor, to enable comparison of the extracted parameters. The factors were stress, vowel quality, and the position of the individual gesture in the word. Gestures were identified as being either stressed or unstressed, and the words contained either /a/ or /i/. The position of the gesture in the utterance was identified by the syllable with which it was associated and by its direction of movement (opening or closing). Differences associated with this factor might be related to final lengthening, which was expected to lower the frequency for gestures in the final syllable.

The articulatory movements studied were the vertical movements of the lower lip in space. One female speaker of American English was recorded, using a Selspot system with LEDs on the nose, upper lip, lower lip and chin. The speaker produced four utterances containing contrasts among the linguistic factors discussed above: ['bibəbib], ['babəbab], [bibə'bib], and [babə'bab]. The utterances were produced in the carrier phrase "It's a __ again." Eleven tokens of each were collected, except for the first utterance, for which fourteen tokens were collected. The data were recorded on an FM tape recorder, then sampled at 200 Hz. The trajectories were smoothed using a 25 ms triangular window.

For each token, the data curve representing the trajectory of the lower lip was divided into windows corresponding to the opening and closing gestures. Each window was marked in two ways, Peak and CV, as described in section 2.1. The edges of the CV windows were marked at points on the movement trajectory where the displacement from a movement extremum exceeded 0.58 mm. A sample utterance with the six windows marked is shown in Figure 6(a), Peak windows, and Figure 6(b), CV windows.

Analyses of the data were performed with the following factors in the statistical analyses: Syllable in the utterance (1, 2, or 3), Direction of movement (Opening or Closing), Stress (Stressed or Unstressed), Window Type (Peak or CV), and quality of the full Vowel in the utterance (/i/ or /a/). Figures 6(a) and (b) show how the gestures were associated with different values of the linguistic factors, for Peak and CV windows respectively. Within each syllable, gestures were associated with a Direction of movement: either opening (lowering the lip and jaw for a vowel) or closing (raising to form a bilabial closure). Each syllable consisted of one gesture in each direction. Thus the interaction of Direction and Syllable picks out individual gestures in the utterance.

In each utterance, stress fell on either the first full vowel (the first syllable) or the second full vowel (in the final syllable). The medial syllable was always reduced. If stress fell on the first vowel, gestures 1 through 3 (the opening into the first vowel through the opening into the schwa) were categorized as stressed, and gestures 4 through 6 (the closing out of the schwa through the closing out of the second full vowel) as unstressed. If stress was on the second full vowel, the first group of gestures was considered unstressed and the second stressed.

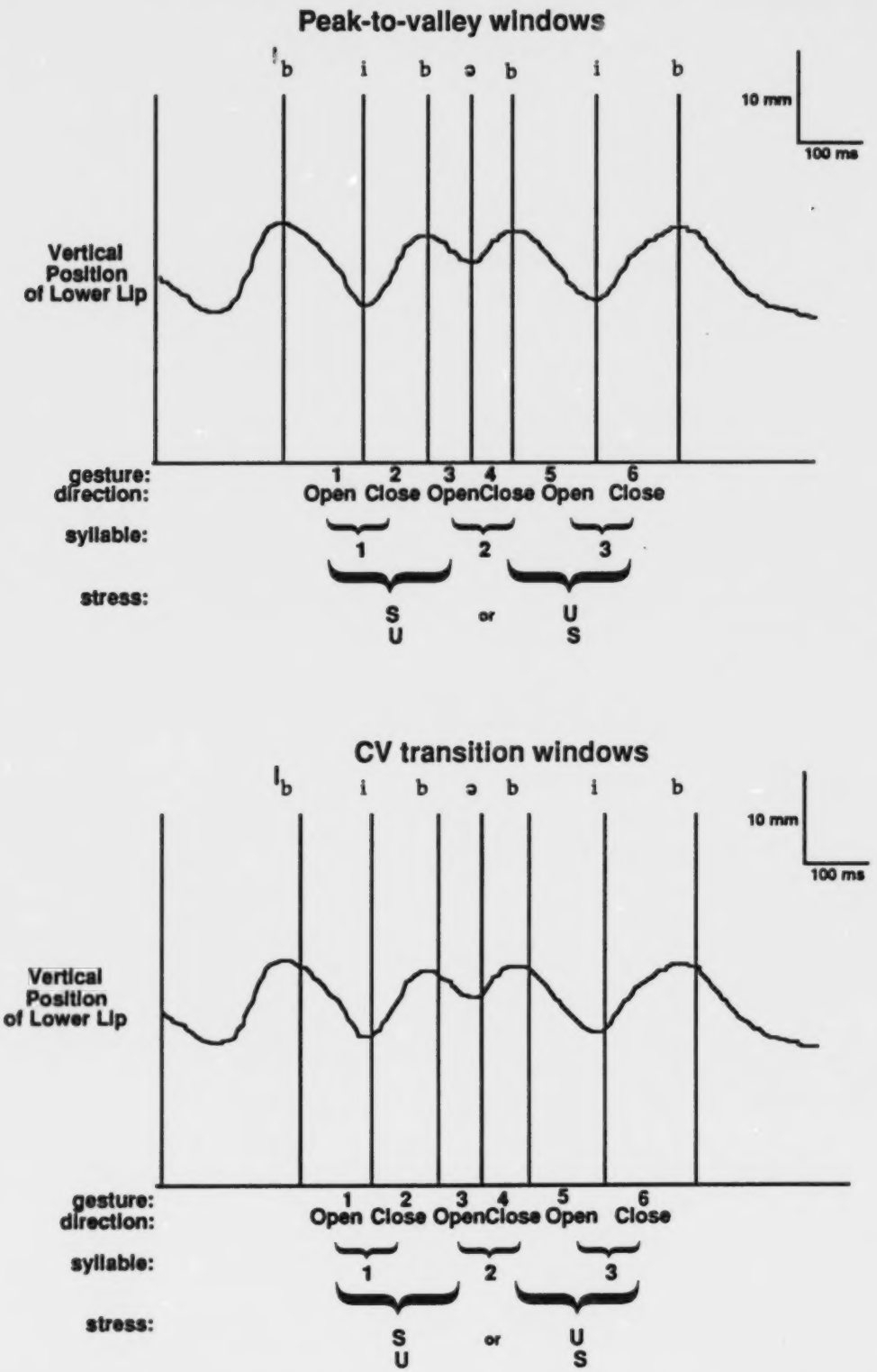


Figure 6. **A sample token with the assignment of the values of the linguistic factors shown below.** *(a)* **Peak windows,** *(b)* **CV windows.**

Thus gestures for the medial schwa were categorized as stressed or unstressed depending on the neighboring syllable. This grouping, while not immediately obvious, was chosen because results from preliminary analyses of these data had shown that the stress effects for the two gestures for the schwa tended to pattern with their adjacent full vowel, not with each other. In the interest of reducing the complexity of the interactions, we decided to divide the stressed and unstressed portions of the utterance between the two gestures of the middle reduced syllable. In a similar fashion, all gestures in an utterance in which the full vowels were /i/ were categorized with /i/, even though the gestures for the schwa were, of course, not directly involved in the production of /i/.

We were mainly interested in the effects of the linguistic factors and Window Type on natural frequency. The least error fits were analyzed first, since they are the most accurate and should therefore provide the most accurate values for frequency. Then, the damping ratio was held constant at three different values (.2, .5, and .8). The values for frequency that were provided at these three different damping ratios were compared to the values provided by the least error fits to see how much accuracy was lost by using a fixed damping ratio rather than allowing it to vary, as was done in the least error fits.

No analysis was done of the constant level parameter values. When the "target" of a movement is unknown, at any damping ratio less than critical damping the constant level becomes increasingly context-sensitive as the damping ratio lowers. In order to confirm that the movements showed the expected patterns in amplitude changes, measurements of the amplitude of the movements were made using the peak-to-valley windows. They showed that, as expected, stressed gestures had larger amplitude movements than unstressed (mean of 7.53 mm versus 5.45 mm). Full vowels had larger amplitude than schwa (8.52 versus 2.31 mm). The stressed gesture with the largest mean amplitude was the opening into the first full vowel (11.72 mm). The gesture with the smallest amplitude was the opening into the schwa when it followed an unstressed vowel (2.24 mm).

3.2 Analysis of least error fits

PARFIT fits were obtained for all windows of all data files, using both Peak and CV windows, at damping ratios from 0.0 to 1.0, using the Boundary condition at damping ratios of 0.0 to 0.8, and Initial condition above that, which were the conditions that had been shown to be preferred by the tests with simulated data. Each window in the articulatory data was assumed to correspond to the beginning of a new regime. The fit that had the least error was selected for each window of data and window type (Peak and CV). This means that within the same data file, the "least error fits" for different windows could have different damping ratios.

Analyses of variance using BMDP 4V were run on the extracted values for frequency and the damping ratios that gave the fits with the least error. Effects with p values below .05 were considered significant. Where main effects and interactions were significant, simple main effects were run to test the extent to which the significance held up in all conditions. In certain cases, post-hoc Newman-Keuls tests were also used.

All the gestures were fit quite accurately by the damping ratio with the least error. In calculating the mean squared error, the amplitude of the data was normalized to be between -1 and 1. Although the size of error of these fits varied among the gestures, the mean errors for all the gestures fell in the same order of magnitude. The gesture with the smallest error was the closing out of the first full vowel (syllable 1), with mean squared error of .00012. The largest was the opening into the first full vowel (syllable 1), with a mean of .00059. Overall, opening gestures had larger mean squared error than closing gestures, .00040 for opening and .00019 for closing. This difference was significant overall in an analysis of variance, ($F=78.43$, $p<.001$), but was also involved in numerous interactions.

3.2.1 Natural frequency and phase angle

3.2.1.1 Results

Frequency. The extracted values for natural frequency, which were constrained to be between 0 and 20 Hz (see section 2.2), ranged in the least error fits from 1.84 Hz (for a stressed opening gesture into the final full vowel, in syllable 3) to 12.56 Hz (for an unstressed opening gesture into the schwa, in syllable 2). For natural frequency, the main effects of Syllable, Direction, Stress, and Window Type were significant overall, but each interacted with other factors. These will be examined individually. The main effect of Vowel did not reach significance. The F-values, degrees of freedom, and significance levels for this analysis are given in Table 3.

Table 3. *F-values, degrees of freedom, and significance for analyses of variance of natural frequency of fits with least error (variable damping ratios), and of fits at 3 fixed damping ratios. All 2- and 3- way interactions are shown; only those 4-way interactions that reached significance in at least one analysis are shown.* *** *indicates significance of* $p<.001$, ** *of* $p<.01$, *and* * *of* $p<.05$.

		Damping ratio			
	df	Least error	.2	.5	.8
Syllable	2,516	132.58***	99.60***	205.55***	–
Direction	1,516	353.20***	106.54***	162.87***	24.10***
Stress	1,516	163.08***	73.85***	110.05***	21.22***
Vowel	1,516	–	–	–	–
Window Type	1,516	159.60***	182.12***	452.06***	63.93***
Syllable × Direction (= Gesture)	2,516	115.60***	108.49***	95.61***	17.57***
Syllable × Stress	2,516	26.28***	13.55***	21.86***	–
Syllable × Vowel	2,516	–	–	5.11**	–
Syllable × Window Type	2,516	–	6.43**	6.70**	9.53***
Direction × Stress	1,516	4.88*	–	–	–
Direction × Vowel	1,516	–	4.04*	6.37*	–
Direction × Window Type	1,516	125.16***	145.85***	–	–
Stress × Vowel	1,516	7.40**	–	11.70***	–
Stress × Window Type	1,516	7.25**	–	6.09*	–
Vowel × Window Type	1,516	–	–	–	–
Syllable × Direction × Stress	2,516	6.09**	–	–	–
Syllable × Direction X Vowel	2,516	16.06***	5.32**	8.86***	–
Syllable × Direction × Window Type	2,516	64.56***	55.26***	39.99***	55.12***
Syllable × Stress v Vowel	2,516	–	–	3.02*	–
Syllable × Stress × Window Type	2,516	–	–	–	4.09*
Syllable × Vowel × Window Type	2,516	–	–	–	–
Direction × Stress × Vowel	1,516	–	–	4.51*	–
Direction × Stress × Window Type	1,516	–	14.03***	–	–
Direction × Vowel × Window Type	1,516	–	–	–	–
Stress × Vowel × Window Type	1,516	–	–	–	11.49***
Syllable × Direction × Stress × Vowel	2,516	–	–	3.90*	–
Syllable × Direction × Stress × Window Type	2,516	–	–	–	6.46**
Syllable × Direction × Stress × Vowel × Window Type	2,516	–	–	–	4.82**

In addition to the effect of Syllable, the interactions Syllable × Direction, Syllable × Stress, Syllable × Direction × Stress, Syllable × Direction × Vowel, and Syllable × Direction × Window were significant. Simple main effects showed that the main effect of Syllable on natural frequency was significant everywhere, and post-hoc Newman-Keuls tests showed that the natural frequency for each syllable was significantly different from that of each other syllable, with Syllable 2 having the highest frequency, Syllable 3 the lowest, and Syllable 1 an intermediate value. These values are plotted in the leftmost group in Figure 7.

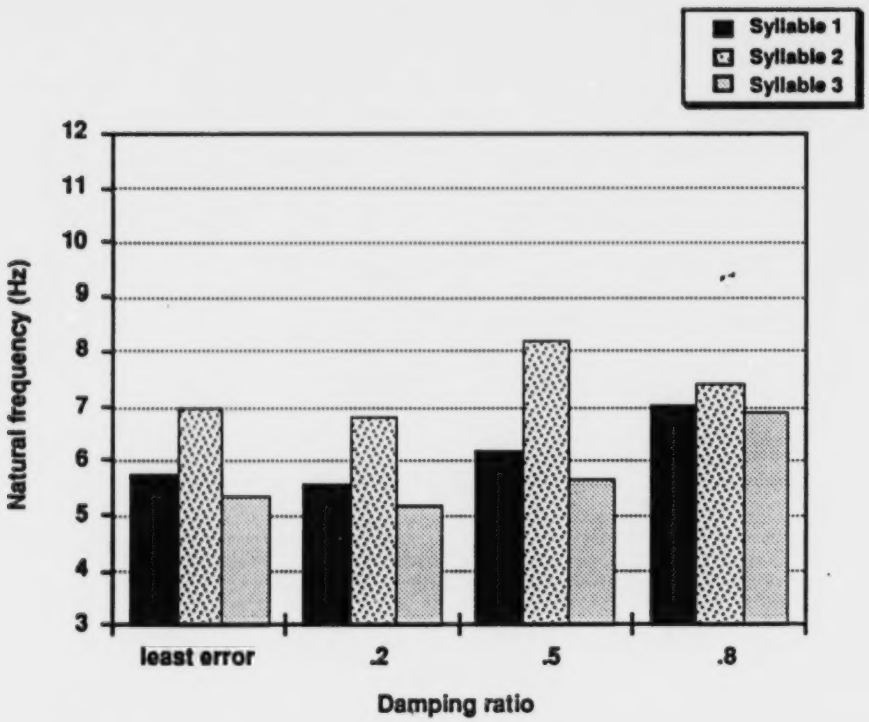


Figure 7. Natural frequency values for each syllable in the least error fits and in fits at fixed damping ratios of .2, .5 and .8 for articulatory data.

The main effect of Direction (opening/closing) on natural frequency was significant overall, and the following interactions were significant: Direction × Syllable, Direction × Stress, Direction × Window, Direction × Syllable × Stress, Direction × Syllable × Vowel, and Direction × Syllable × Window. However, simple main effects showed that because of the interaction of Syllable and Direction, the significance of Direction held only in the syllables with full vowels (syllables 1 and 3). In these syllables, the frequency of closing gestures was significantly higher than of opening gestures, as can be seen in Figure 8. The difference between the two directions was not significant in the reduced middle syllable.

The interaction of Syllable × Direction, which corresponds to the effect of individual gestures, was also significant everywhere. Newman-Keuls tests showed that significant differences existed between three groups of these gestures, plotted in Figure 8. The lowest frequency gestures, labeled Group A, were the opening gestures for the two full vowels. Group B, with frequency significantly higher than Group A, is the closing gesture of the final syllable. The remaining gestures, Group C, generally associated with the schwa, had frequency values significantly higher than Groups A and B. This means, then, that the frequencies of the opening gestures for the full vowels (syllables 1 and 3) were significantly lower than the frequency of the opening gesture of the schwa (syllable 2), while the final closing gesture (syllable 3) has a lower frequency than the other closing gestures.

Stress was also significant overall, with the following interactions: Stress × Syllable, Stress × Direction, Stress × Vowel, Stress × Window, and Stress × Syllable × Direction. The main effect of Stress on natural frequency was significant in the first two syllables. In Figure 9 it can be seen that all stressed syllables had lower frequency values than the corresponding unstressed syllables, but in the final syllable this difference was so small that it was not significant statistically. This result is consistent with the results of acoustic experiments showing that final unstressed syllables can be as long as final stressed ones (Oller, 1973).

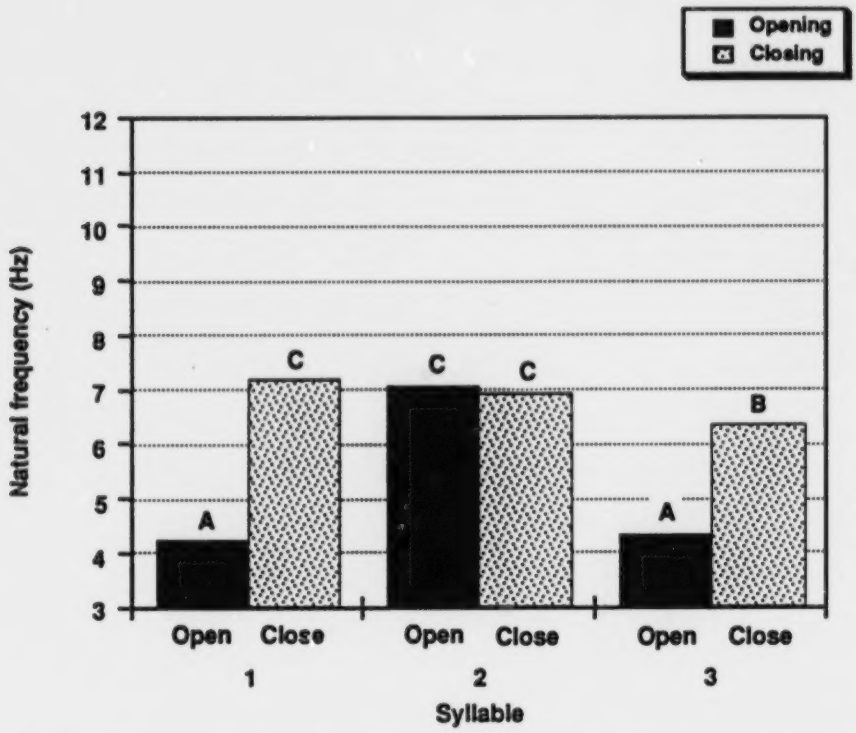


Figure 8. Natural frequency values for the individual gestures of the least error fits for articulatory data. The letters indicate gestures that could be grouped by their significant differences (using Newman-Keuls tests).

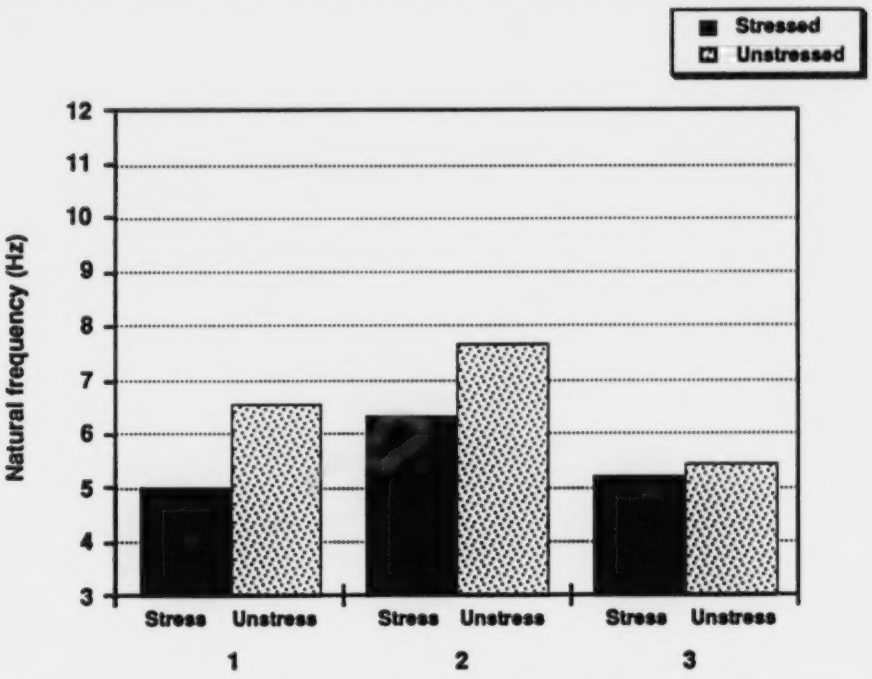


Figure 9. Natural frequency values in the least error fits for articulatory data compared between stressed and unstressed conditions for the three syllables.

Recall that in syllable 2, where the vowel was a schwa, stress was assigned to individual gestures, not to the syllable as a whole. The opening gesture in syllable 2 was considered stressed when it followed a stressed vowel, and the closing gesture was considered stressed when it preceded a stressed vowel. Thus the gestures in syllable 2 were never both stressed or both unstressed within the same utterance. As can be seen in Figure 10, the effect of Stress (higher frequency for unstressed) was much greater in the closing gesture of syllable 2 than in the opening gesture of the same syllable. However, all gestures had higher frequency when unstressed, even where the effect of Stress was not significant, as in syllable 3. There was also a significant interaction of Stress × Vowel, such that the tendency of unstressed gestures to have higher frequencies than stressed gestures was amplified in utterances with the vowel /a/ compared to utterances with /i/.

The effect of Window Type on natural frequency, which was significant overall, was involved in the interactions Window × Direction, Window × Stress, and Window × Direction × Syllable. The main effect of Window Type was found, by using simple main effects, to be significant only in opening gestures. Higher frequency values were obtained using CV windows in both opening and closing gestures, but the difference between the two window types was significant only in opening gestures. These values are plotted in Figure 11.

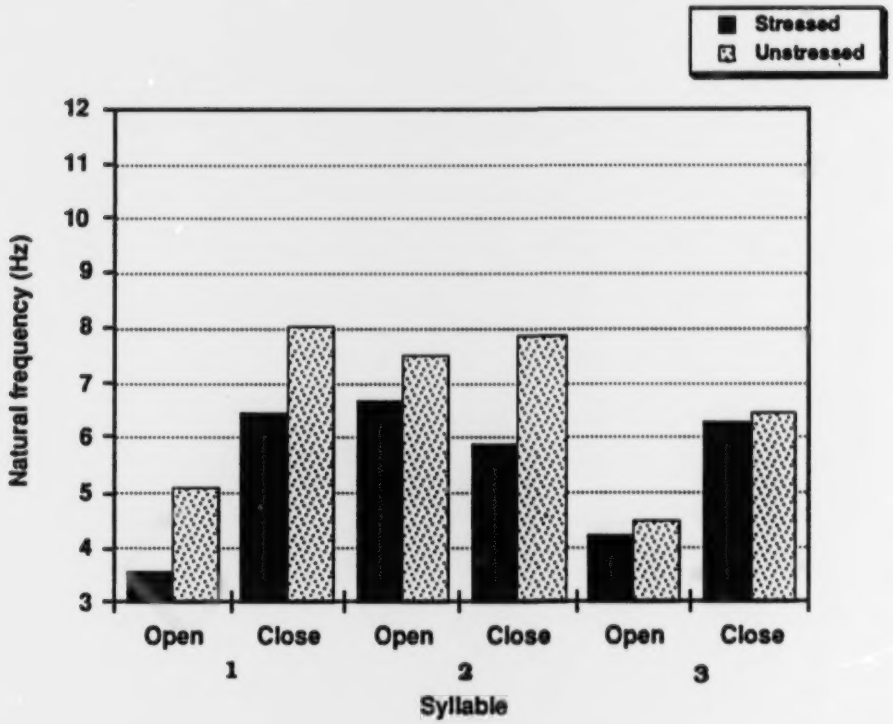


Figure 10. Natural frequency values in the least error fits for articulatory data compared between stressed and unstressed conditions in the individual gestures.

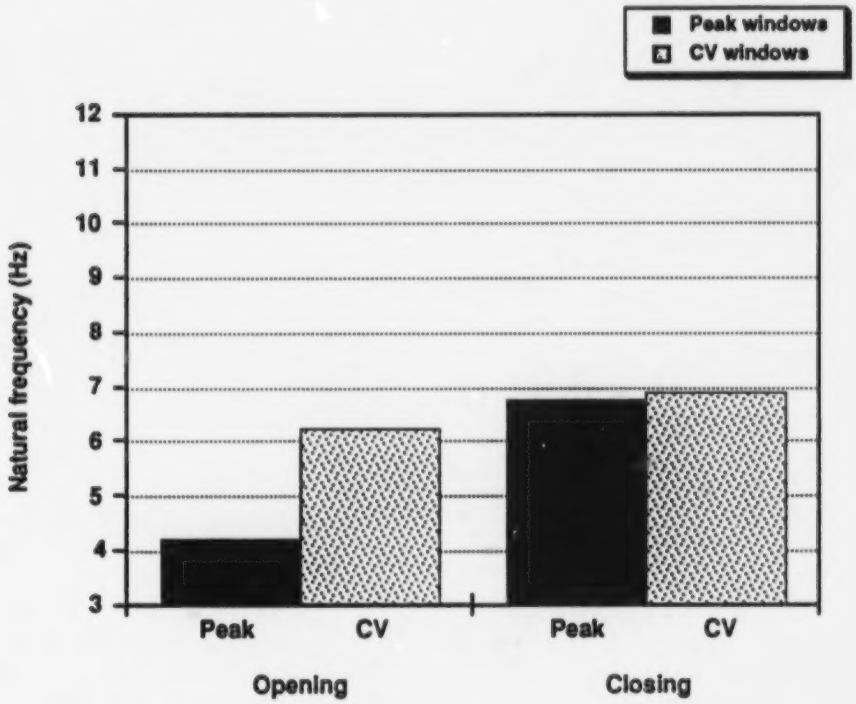


Figure 11. Natural frequency values in the least error fits for articulatory data, showing the effect of Window Type in opening and closing gestures.

Although the interaction of Window × Direction × Syllable did not change the significance of the main effect of Window, it did affect some of the patterns found among individual gestures. Recall that the overall pattern was for opening gestures to have lower frequencies than the closing gestures within the same syllable. This effect was changed by Window Type in the middle syllable (see Figure 12). When analyzed with Peak windows, as expected the frequency of the opening gesture (5.44 Hz) was lower than of the closing gesture (7.58 Hz). However, the reverse was true when this middle syllable was analyzed with CV windows. The closing gesture for the middle syllable was also unique in having a higher frequency with Peak windows than with CV windows.

Phase angle. Phase angle was investigated primarily as a further confirmation that the analyses were sensible, rather than as a research question. In general, the relation between natural frequency and phase angle was that expected (see section 2.2). First, the average phase angle was 187 degrees, which means that the movements were being analyzed as being approximately half a cycle, as intended. Table 4 lists the phase angles by Syllable position, Direction, and Window Type. Second, there was a strong tendency for phase angle to be larger when the frequency was higher. This was true in particular for Stress, Direction, and Window Type, all of which had significant effects on phase angle as well as on frequency; moreover, as for frequency, Vowel did not have a significant effect on phase angle. The effect of Syllable position on phase angle, however, was the reverse of its effect on frequency in those limited environments in which it was significant.

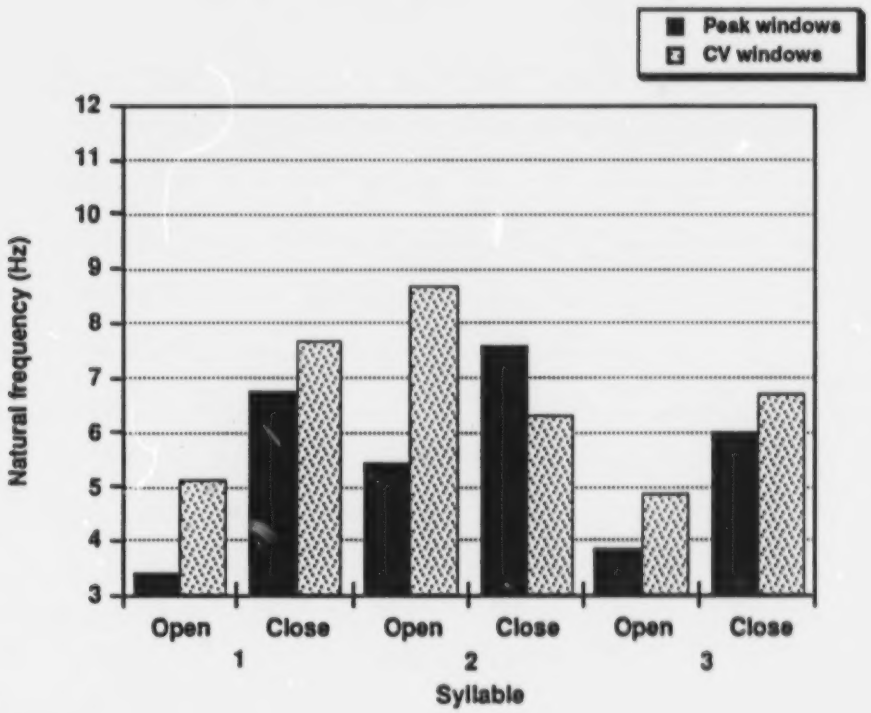


Figure 12. **Natural frequency values in the least error fits for articulatory data, showing the effect of Window Type in the individual gestures.**

Table 4. *Phase angles.*

	CV windows		Peak windows	
Syllable	Closing	Opening	Closing	Opening
1	238	191	181	145
2	182	208	202	143
3	226	186	187	157

Since a number of 2-way and 3-way interactions were significant for all the main effects of phase angle, simple main effects were performed using the same strategy as for the frequency, with the following results. Stress had a significant effect on phase angle in syllables 1 and 2 (but not in syllable 3), and in closing gestures using CV windows; on the average, unstressed movements were analyzed as having larger phase angles than stressed movements (unstressed versus stressed: in syllables 1 and 2, 194 versus 179 degrees; in closing gestures using CV windows, 229 versus 201 degrees). Direction had a significant effect on phase angle everywhere except in the stressed CV condition and except in the second syllable of words containing /i/; on the average, closing gestures had larger phase angles than opening gestures (203 versus 172 degrees). The effect of Window Type on phase angle was significant everywhere; on the average, CV windows had larger phase angles than Peak windows (205 versus 169 degrees). The effect of Syllable position was significant only in restricted environments: for stressed gestures, for closing gestures in words with /i/, and for opening gestures in words with /a/. In these environments, it showed the reverse order for phase angle (syllable 3 >= syllable 1 >= syllable 2) as for frequency: for stressed gestures, 188 versus 181 versus 176; for closing gestures in words with /i/, 214 versus 214 versus 191; and for opening gestures in words with /a/, 183 versus 174 versus 163.

3.2.1.2 Discussion

The values for natural frequency provided by the least error fits were expected to show patterns in the linguistic factors that reflect the kinematic properties of the utterances. For example, since the gestures associated with the reduced syllable

(syllable 2) have a shorter duration and smaller displacement than gestures in other syllables, they were expected to have a higher frequency than those associated with the full vowel syllables (syllables 1 and 3), as was indeed the case. Moreover, the final syllable had the lowest frequency. This could be a consequence either of lengthening at the end of the utterance, which would tend to lower the frequency of the final syllable, or a consequence of the closing gesture in the first syllable closing into a reduced syllable, thereby increasing the frequency of the first syllable.

The analysis of the effects of individual gestures (the interaction Syllable × Direction) suggested that both sources of syllable differences may be important. Not only the two gestures associated with syllable 2, but also the closing gesture of syllable 1, immediately preceding the reduced syllable, had significantly higher frequencies than the other gestures. As might be expected from an articulatory perspective, the movement into the reduced syllable grouped with the patterns of the movements within the syllable. The high frequency of the closing gesture of syllable 1 is also at least partly attributable to the overall higher frequency of the closing gestures in the full vowel syllables compared to the frequency of the opening gestures of those syllables, a result that is consistent with the steeper Peak velocity/Displacement slopes that Kelso et al. (1985) and Vatikiotis-Bateson (1988) found to be characteristic of closing gestures compared to opening gestures.

Significantly lower frequencies were obtained in the stressed condition than in the unstressed condition in the first two syllables, as was expected from the results obtained by Browman and Goldstein (1985), Kelso et al. (1985), and Vatikiotis-Bateson (1988). In the final syllable, however, while the effect of Stress was still to lower the frequency in the stressed condition as compared to unstressed, the difference between stressed and unstressed was not significant, probably because the frequency of the final unstressed syllable was also lower (compared to the other unstressed syllables). As noted above, this is consistent with final lengthening.

The patterns for phase angle were generally similar to the patterns for natural frequency, with those conditions analyzed as having higher frequency also analyzed as having a larger phase angle. Thus, unstressed gestures in non-final syllables were generally analyzed as consisting of a larger portion of a higher frequency cycle than stressed gestures in non-final syllables. Similarly, closing gestures were analyzed as being a larger portion of a higher frequency cycle than the opening gesture in the same syllable (at least for syllables with full vowels). And CV windows resulted in analyses showing larger portions of higher frequency cycles than Peak windows (at least for opening movements). While there were trends in Syllable position in which the higher frequency syllables were analyzed as consisting of larger portions of the cycle, in the limited environments in which phase angle significantly differed there was an inverse relation between frequency and phase angle.

Looking at how the patterns for natural frequency were affected by Window Type, in general they were the same regardless of the choice of Window Type, with two exceptions. First, absolute values of both frequencies and phase angles differed between the two Window Types in that analyses using CV windows gave larger portions of higher frequency curves than analyses using Peak windows. (The limitation of significance to opening gestures for the frequency effect was likely due entirely to the reversal in the closing gesture of the reduced syllable 2 observable in Figure 12. We will argue below that this gesture is generally anomalous.) Since nothing about the curve itself had changed between the CV and Peak analyses, but rather only the particular section of the curve being considered, presumably the primary difference between the two windows was in the phase angle or portion of the cycle, with the natural frequency difference being effectively an artifact of the analysis procedure, given the difference in phase angle.

The second way in which frequency was affected by Window Type was the relative frequencies of opening and closing gestures in the reduced syllable 2. Although in general closing gestures were analyzed as being larger portions of higher frequency cycles, the frequency effect was reversed when the reduced syllable was analyzed using CV windows (see Figure 12). Put another way, in the reduced closing gesture the relation between Peak and CV frequencies was reversed from the relation for the other gestures. This reversal of frequency may be the cause of the limitation of significance for the effect on frequency both of Direction and of Window Type.

Symmetry considerations suggest that the anomaly probably lies in the CV analysis of the reduced closing gesture returning a lower frequency than expected. Thus, in Figure 12, if the general open/close and peak/CV patterns were to apply to the closing gesture in the reduced syllable

2 when analyzed with a CV window, this gesture would be the highest frequency gesture in the figure. Also, in Table 4, the phase angles for the CV closing gestures would have the same pattern as the rest of the table if the phase angle for syllable 2 were equal to or higher than the phase angle of syllable 1, rather than lower as it is. In the discussion of the next section, we will attempt to explain why this gesture might be anomalous. For now, we will simply reiterate that this behavior means that Window Type can affect the results in the frequency analysis. In syllables with full vowels, Window Type affects only the absolute values of frequencies, but not the pattern of any of the results. However, in reduced syllables, the choice of Window Type can have a real impact on the results, going so far as to reverse the directionality of the pattern.

3.2.2 Damping ratio in the least error fits

The results above showed that the natural frequency values extracted by PARFIT do, in the best case (the least error fits) reflect the effects of factors such as stress systematically and in correspondence with previously established results. Since our tests with the simulated data had suggested that the least error fits could be used to estimate the true damping ratio within +/- .1, we were interested to see if there were any patterns of the damping ratios of the least error fits analogous to those of the frequency values associated with them. Recall that, since each window was assumed to be the beginning of a new regime, a single data file (6 windows) could be fit with as many as 6 different damping ratios. Thus the choices of damping ratio that were selected by the least error could show the effects of the linguistic factors and of the type of window.

3.2.2.1 Results

The mean damping ratio for all the least error fits was .13, with values ranging from .0 to .57 across the categories defined by the factors. The factors had less systematic effects on the damping ratios than on the extracted values for natural frequency. The F-values, degrees of freedom, and significance levels for the analysis of variance of damping ratio are listed in Table 5. As in natural frequency, the main effects of Syllable, Direction, and Stress were significant overall. Unlike the natural frequency results, for damping ratio the main effect of Vowel was significant overall, but Window Type was not. Since all the main effects were involved in interactions, simple main effects analyses were run, which showed that in fact all of them were significant only in limited environments.

Table 5. *F-values, degrees of freedom, and significance for analysis of variance of damping ratio for fits with least error (variable damping ratios). All 2- and 3- way interactions are shown; only the single 4-way interaction that reached significance is shown.* *** *indicates significance of p<.001,* ** *of p<.01, and* * *of p<.05.*

	df	Least error
Syllable	2,516	47.77***
Direction	1,516	210.21***
Stress	1,516	7.72**
Vowel	1,516	13.13***
Window Type	1,516	0.58
Syllable × Direction	2,516	33.51***
Syllable × Stress	2,516	–
Syllable × Vowel	2,516	–
Syllable × Window Type	2,516	52.03***
Direction × Stress	1,516	12.68***
Direction × Vowel	1,516	4.27*
Direction × Window Type	1,516	14.92***
Stress × Vowel	1,516	–
Stress × Window Type	1,516	22.88***
Vowel × Window Type	1,516	–
Syllable × Direction × Stress	2,516	–
Syllable × Direction × Vowel	2,516	–
Syllable × Direction × Window Type	2,516	59.18***
Syllable × Stress × Vowel	2,516	–
Syllable × Stress × Window Type	2,516	4.19*
Syllable × Vowel × Window Type	2,516	–
Direction × Stress × Vowel	1,516	–
Direction × Stress × Window Type	1,516	17.57***
Direction × Vowel × Window Type	1,516	–
Stress × Vowel × Window Type	1,516	8.82**
Syllable × Direction × Vowel × Window Type	2,516	6.60**

Syllable was involved in the interactions of Syllable × Direction, Syllable × Window, Syllable × Direction × Window, Syllable × Stress × Window, and Syllable × Direction × Vowel × Window. Because of these interactions, the effect of Syllable on the damping ratio was significant only in closing gestures with Peak windows. This can be seen in Figure 13, in which it is apparent that the significance was due primarily to the closing gesture of the middle reduced syllable, which had a much higher damping ratio than the closing gestures of the other syllables. However, the opening gestures analyzed with Peak windows all had approximately the same damping ratio. With CV windows, the different syllables had similar mean damping ratios.

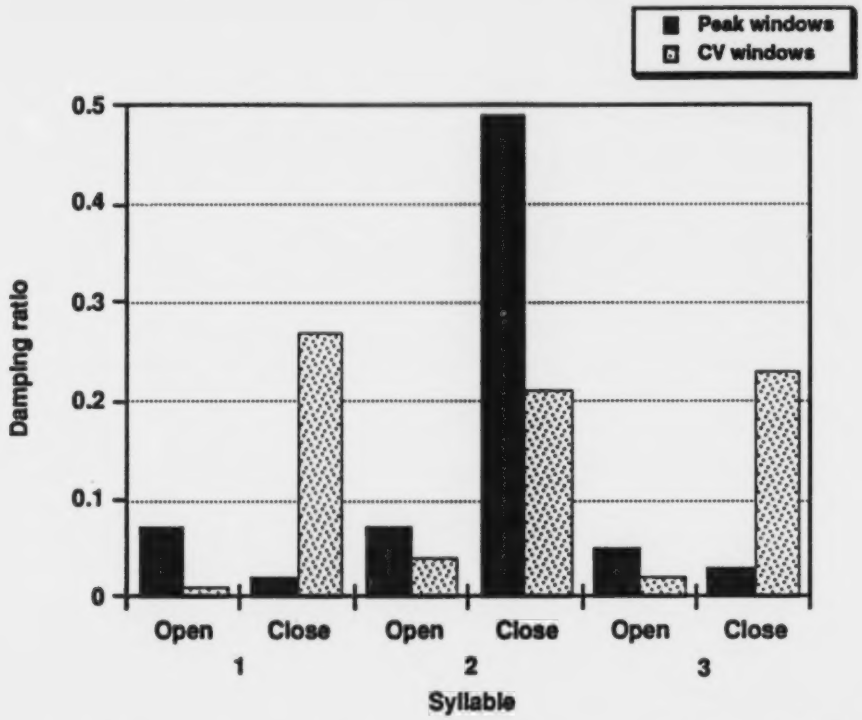


Figure 13. **Damping ratio in the least error fits for articulatory data, showing the effect of Window Type in the individual gestures.**

Direction had the following interactions: Direction × Syllable, Direction × Stress, Direction × Vowel, Direction × Window, Direction × Syllable × Window, Direction × Stress × Window, and Direction × Syllable × Vowel × Window. The main effect of Direction on damping ratio was significant only in CV windows, and with Peak windows in syllable 2 (the reduced syllable). This can be seen in Figure 13, where the closing gestures can be seen to have higher damping ratios than the corresponding opening gestures in these instances. There was no significant difference, with Peak windows, in the damping ratios of the opening and closing gestures in the full vowel syllables (syllables 1 and 3).

Simple main effects showed that the interaction of Syllable × Direction, which compares individual gestures, was significant in Peak but not CV windows. Most of the variation in CV windows resulted from the effect of Direction, so the interaction with Syllable was not significant, but in Peak windows the reduced middle syllable showed a different pattern than the syllables with full vowels, resulting in a significant interaction between Syllable and Direction. Post-hoc Newman-Keuls tests showed that, in Peak windows, the damping ratio of the closing gesture of the middle syllable was significantly higher than the damping ratio of any other gesture; the other gestures did not differ significantly from one another. This can also be seen in Figure 13.

Because of the interactions of Stress × Direction, Stress × Window, Stress × Syllable × Window, Stress × Direction × Window, and Stress × Vowel × Window, the main effect of Stress on damping ratio was significant only in closing gestures analyzed with CV windows. As can be seen in Figure 14, in this case alone, unstressed gestures had much higher damping ratios than stressed gestures. However, elsewhere Stress did not significantly affect the damping ratio.

The main effect of Vowel was shown, using simple main effects, to be significant only in limited instances. Its interactions were Vowel × Direction, Vowel × Stress × Window, and Vowel × Syllable × Direction × Window. Words with the vowel /i/ had significantly higher damping ratio than words with /a/ in stressed gestures with CV windows and in unstressed ones with Peak windows. This effect can be seen in Figure 15. The effect of vowel quality was also significant in the closing gestures of the first and last syllables when analyzed with CV windows, where gestures in words with /i/ (which has smaller amplitude movement) had significantly higher damping ratios than in the corresponding words with /a/.

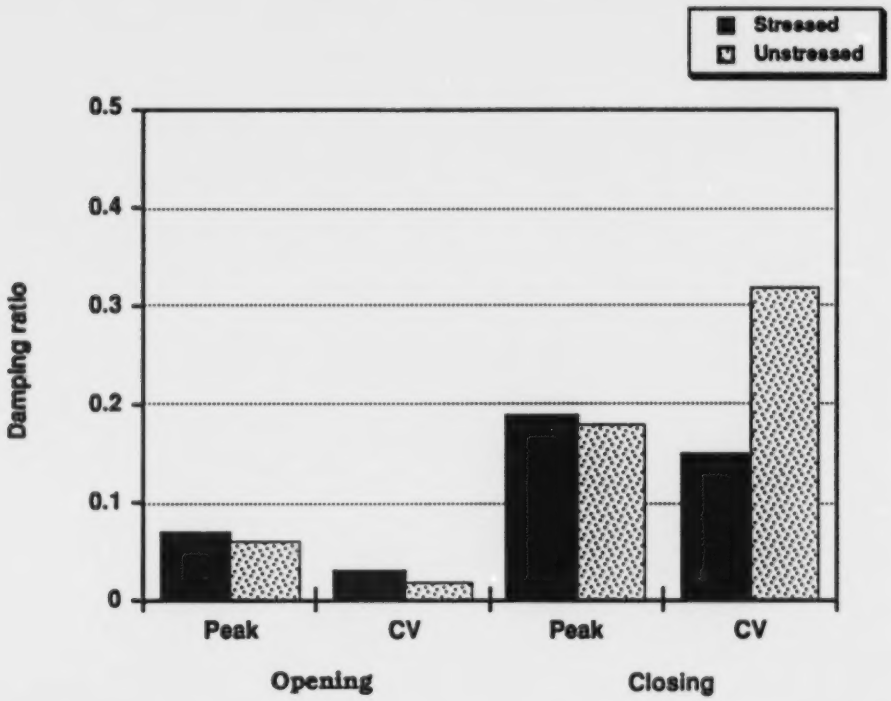


Figure 14. **Damping ratio in the least error fits for articulatory data compared between stressed and unstressed conditions for the different Window Types in opening and closing gestures.**

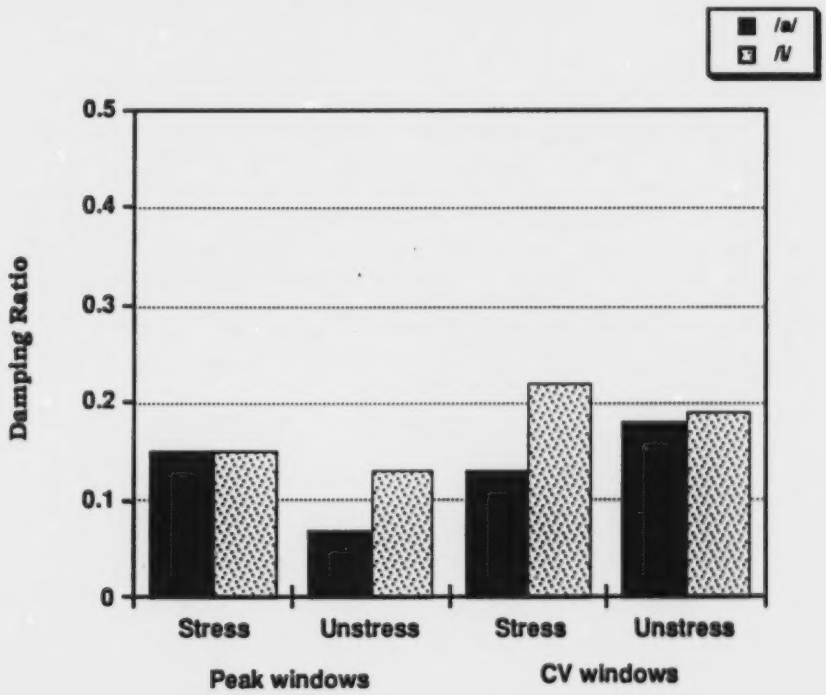


Figure 15. **Damping ratio in the least error fits for articulatory data compared between /a/ and /I/ in stressed and unstressed conditions for the different Window Types.**

3.2.2.2 Discussion

Assuming that the fixed damping ratios that gave the least error for the articulatory data gave estimates of the true damping ratio of the data that were approximately as accurate as those of the simulated data, the analyses of these damping ratios showed that, in general, neither the linguistic factors nor the Window Type had much effect on the damping ratio. Much of the time there simply was little difference among different environments, with the damping ratio averaging around 0.1. Moreover, the few significant differences were found in much more limited environments than those that held for natural frequency.

In particular, unlike the case for natural frequency, Stress had no systematic effect on damping ratio, and neither did Window Type, while Vowel had some effect on damping ratio but none on frequency. Moreover, the very limited Syllable effect on damping ratio presumably was due solely to a single gesture, the closing gesture in the reduced syllable 2, which had a much higher damping ratio (.49) when analyzed using Peak windows. Direction was the only factor to affect both damping ratio and frequency.

These results can be related to the general characteristics of the shape of the observed articulatory curve being analyzed. In so doing, it appears that the results may indeed be suggestive of underlying causes. Consider first the effect of Window Type. The CV window contained the curved, or flattened, portion at the end of the movement, while the Peak window contained only a bit of this curved portion. Since a curve that flattens out as it approaches the end of the movement can be associated with a higher damping ratio (see Figure 1), the flat portion at the end of the CV windows might have been analyzed as a higher damping ratio, but it was not. There was no significant difference in damping ratio between the Window Types. Rather, the difference between Window Types resided in the CV windows being analyzed as being a larger portion of the curve than the Peak windows, as evidenced by phase, with the curve additionally being significantly higher frequency (at least in opening gestures).

Stress also did not affect damping ratio, but only the frequency. That is, the larger amplitude, slower stressed movements were analyzed as being a smaller portion of a lower frequency curve than unstressed movements (when they were non-final), but not as differing in damping ratio. The only exception was in the comparatively highly damped closing gestures, when analyzed using CV windows that saw the entire compression: in these gestures, unstressed movements were analyzed as being significantly more damped than stressed movements, as well as higher frequency.

In general, with one exception (discussed below), damping ratio differences were analyzed as occurring only in those situations in which they might reasonably be expected to occur on physical grounds. One situation in which damping ratio was analyzed as being increased was for the vowel with relatively little movement (/i/), which was sporadically analyzed as having a higher damping ratio than the vowel with relatively great movement (/a/). The second situation was the increased damping ratio for closing gestures (relative to opening gestures) when analyzed using CV windows (the effect of Direction). That is, the damping ratio was analyzed as being increased when the window of analysis included a flatter portion at the end of the window such as that observed during the probable compression of the lips during bilabial closure. The CV Window Type presumably contributed to the detection of this effect by including more of the compressed portion of the curve. However, the inclusion of more of the flatter portion of the curve did not significantly change the effects of the linguistic factors on the frequency analyses. To recapitulate, then, in addition to being more highly damped in CV windows, closing gestures were analyzed as being larger portions of higher frequency cycles, compared to opening gestures, for both Window Types (although the anomalous closing gesture in the reduced syllable 2 limited the significance of the Direction effect on frequency).

Unlike the results obtained for frequency, where Gesture (Syllable × Direction) was significant overall, with three groups of gestures having different frequency values, for damping ratio only one gesture (the closing gesture of the reduced syllable 2) differed significantly from the others, and that only when analyzed with Peak windows. We suggested above that this gesture also caused the limited Syllable position effects on damping ratio that were observed, as well as serving to limit the significance of the frequency analyses for Direction and Window Type. Here we argue that this gesture, which appeared to be anomalous in general, represented the only situation in which an analysis of increased damping ratio probably did not reflect a physical cause. That is, we will argue that in this case, the increased damping ratio is an artifact of the analysis technique rather than an attribute of the data.

The closing gesture in syllable 2 differed from the other two closing gestures in that it alone was followed by a non-reduced syllable. Thus, the peak that it closed into had the characteristics of initial consonants in non-reduced syllables, which included a broader peak than in the reduced syllable (see Figure 6). When a Peak window was used, this broader peak was was apparently interpreted as the flattening associated with increased damping ratio; however, when a CV window was used, the broader peak seems to have been interpreted as the broadening associated with decreased frequency. Such an interpretation is supported by further analysis showing that stress (which generally broadened the peaks) increased the damping ratio for only this gesture, and only when it was analyzed using a Peak window. Note that the phase angle for the closing gesture in the reduced syllable 2, when analyzed using a CV window, was considerably smaller than the phase angles in syllables 1 and 3, unlike the Peak analysis. That is, given the short, small amplitude movement of the closing gesture in syllable 2, the flattish peak occupied a substantial proportion of the overall duration in the CV analysis; in this case the curve within the window was analyzed as a smaller portion of a lower frequency cycle.

Since we know that in general, the non-reduced syllables were significantly lower in frequency than the reduced syllable, but did not differ in damping ratio, the CV analysis appears to be more in tune with the overall results. That is, the CV analysis reflects the effect of the following syllable on the closing gesture into that syllable more than the Peak analysis. However, in both the Peak and CV windows, the curve being investigated in this syllable 2 closing gesture shares attributes of both the reduced syllable 2 (in the moving portion) and the non-reduced syllable 3 (in the extremum portion). Therefore, another type of window might be usefully investigated, perhaps one that, like the CV window, includes an entire position extremum in a single window, but that includes the preceding extremum, rather than the following.

In terms of the windows investigated in the current work, the Peak windows appear to be appropriate for investigations that are not interested in damping ratio. That is, analyses using Peak windows showed highly regular linguistic effects on frequency, and basically no linguistic effects on damping ratio (except for the anomalous gesture), which indeed was close to being undamped. This is very encouraging for the accuracy and validity of those published articulatory analyses that ignore damping ratio and use Peak windows. However, if information on damping is desired, then CV windows, or some variant in which all contiguous extremum information is analyzed in the same window, are better employed, since the CV-type windows appear to allow differences in damping to be picked up that are not apparent in Peak analyses. If CV windows are used, then the phase angle must be sharply watched, since a window that includes information from two underlying control regimes may return anomalous frequency information.

3.3 Patterns in natural frequency using constant damping ratios

It is on occasion desirable to introduce further constraints on the data analysis procedure, either for theoretical or pragmatic reasons. One constraint we wished to explore because of our modelling work was the assumption of a single constant damping ratio throughout. If such an assumption were made, presumably the natural frequency would be estimated less accurately. However, while the frequency values might differ, the effects of the different statistical factors might prove to be robust enough to remain significant even when the damping ratio was held constant. In such a case, even if the extracted parameter values were somewhat inaccurate, comparisons among extracted values would still be valuable if the main statistical effects remained significant. Further tests were done, then, to see how much the main effects changed when the damping ratio was held constant and thus presumably the natural frequency estimated less accurately. That is, we wished to see how much information was lost when a single damping ratio was used to fit all the data, not the different damping ratios that gave the least error in each window.

Three sets of fits were made in which every window of every data file was fit using a single fixed damping ratio. The damping ratios used were .2, .5, and .8, chosen to sample the range 0.0 to 1.0 that had been used to find the least error fits. Since the mean damping ratio overall was .13, the best results for a single fixed damping ratio were expected to occur at .2. All of these fits were made using the Boundary condition. The same statistical procedures were followed for the three fixed damping ratios as for the least error fits.

3.3.1 Results

As the fixed damping ratio increased, so did the mean values obtained for natural frequency. Table 6 shows these values at the different damping

ratios. Not only did the mean frequency increase, but the statistical significance of the factors was somewhat reduced as the damping ratio increased, particularly for the .8 fits. However, the results exhibited the same basic patterns among the frequency values for all the fits. The results of the ANOVAs run at each of the fixed damping ratios are shown in Table 3 in the rightmost three columns. Simple main effects analyses were run for each damping ratio to determine the environments in which the linguistic factors and Window Type significantly affected the estimate of frequency. Table 7 summarizes the results of these analyses, comparing them to the least error analysis.

As summarized in Table 7 and as can be seen in Figure 7, Syllable was significant everywhere in the fits at damping ratios of .2 and .5, as in the least error fits. Likewise, at both of these damping ratios, each syllable's frequency was shown by post-hoc Newman-Keuls tests to be significantly different from each other syllable. Although the fits at damping ratio of .8 showed the same trends as the other damping ratios, the effect of Syllable did not reach significance, because the frequency values for the individual syllables were much closer together than at the lower damping ratios. To summarize, regardless of damping ratio the order of the syllables' frequencies was identical to that found with the least error fits.

Table 6. *Mean frequencies at different damping ratios.*

	damping ratio	mean frequency (in Hz)
least error	mean = .13	6.03
(fixed)	.2	5.85
(fixed)	.5	6.67
(fixed)	.8	7.10

Table 7. *Summary of significance of the main effects for natural frequency, as shown by analyses of variance. The difference between frequency values was in the direction specified, unless a reversal is indicated in the chart by †. Limitations in the extent of significance are listed in the appropriate row. In some analyses the existence of multiple interactions made it necessary to break down the scope of the main effect's significance in more than one way, e.g., Stress by Syllable and by Direction. "Gesture" is identical to the interaction Syllable × Direction, which identifies individual gestures in the utterance.*

√ significant in the environment named at left
— not significant
blank not applicable; significance shown for other environments
† direction of effect reversed

	Least error	.2	.5*	.8
Syllable (3 < 1 < 2)	√	√	√	—
Direction (OPEN < CLOSE)				
Syllable 1	√	√	√	see below
Syllable 2	—	/i/ †	/i/ †	see below
Syllable 3	√	√	√	see below
Peak windows	√	√		Syll 2 × Unstr; Syll 3 × Str
CV windows	√	—		/i/; /a/ × Unstr; /a/ × Str × Syll 1
Gesture (SEE TEXT)	√	√	√	
Stressed				CV
Unstressed				√
Stress (STRESSED < UNSTRESSED)				√
Syllable 1	√	√	√	
Syllable 2	√	√	√	
Syllable 3	—	—	—	
Open	√		/a/	
Close	√		√	
Vowel	—	—	—	—
Window Type (PEAK < CV)			√	
Open	√	√		Syll 2
Close	—	—		Syll 1 × /i/; Syll 1 × Unstr; Syll 2 × /i/ × Unstr† Syll 3 × /i/ × Str

*The interaction Syllable × Direction × Stress × Vowel ($F=3.90$, $p<.05$) was not taken into account in determining the significances for the .5 fits.

As can also be seen in Table 7, the environments in which the main effect of Direction was significant were basically the same for the .2 and .5 fits and the least error fits. In all three fits, Direction was significant in syllables 1 and 3 but not in syllable 2, although the details for syllable 2 differed. In particular, for the .2 and .5 fits, the relation between the opening and closing gestures of the reduced syllable 2 reversed in words with the vowel /i/. The .2 fits also differed slightly in that the Window Type affected the significance of Direction for them, but not for the .5 or least error fits. As with Syllable, the .8 fits showed the same trends for Direction as in the fits for lower damping ratios, but the significance was limited, in this case to only a few environments that were not a simple subset of the environments at the lower damping ratios. Thus, for Direction, the general tendency at all damping ratios was for opening gestures to have a lower frequency than closing gestures.

The interaction Syllable × Direction, which corresponds to the effect of Gesture, was significant everywhere at damping ratios of .2 and .5, as it was in the fits with least error. Post-hoc Newman-Keuls tests were run for each damping ratio to determine which gestures were significantly different from each other. The results of these tests are shown in Figures 16 (.2) and 17 (.5). The pattern of significance for the .2 fits was the same as for the least error fits (illustrated in Figure 8), except that the frequency of the opening gesture for the reduced syllable 2, labeled E in Figure 16, was significantly higher than that of any other gesture, but the frequency of the closing gesture for the same syllable, labeled C, was not significantly different from the frequency of the final closing gesture, labeled B. The pattern found at .5 was also very similar to what was found in the .2 and the least error fits. Each letter in Figure 17 indicates a gesture or group of gestures whose frequencies differ significantly from every other group. However, for .8 fits, Gesture (Syllable × Direction) was not significant everywhere, so no Newman-Keuls tests were made of the differences among individual gestures. Nonetheless the overall pattern among the frequency values for the individual gestures was similar to the other damping ratios. These values are shown in Figure 18, so that they can be compared to the results of the lower damping ratios.

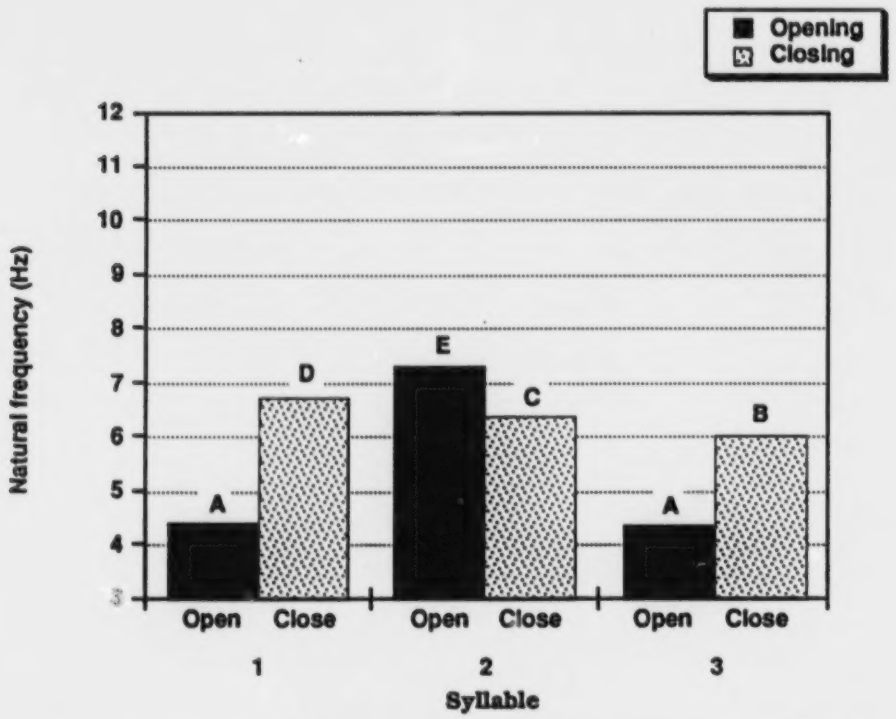


Figure 16. Natural frequency values for the individual gestures of the fits with damping ratio .2. The letters indicate gestures that could be grouped by their significant differences (using Newman-Keuls tests).

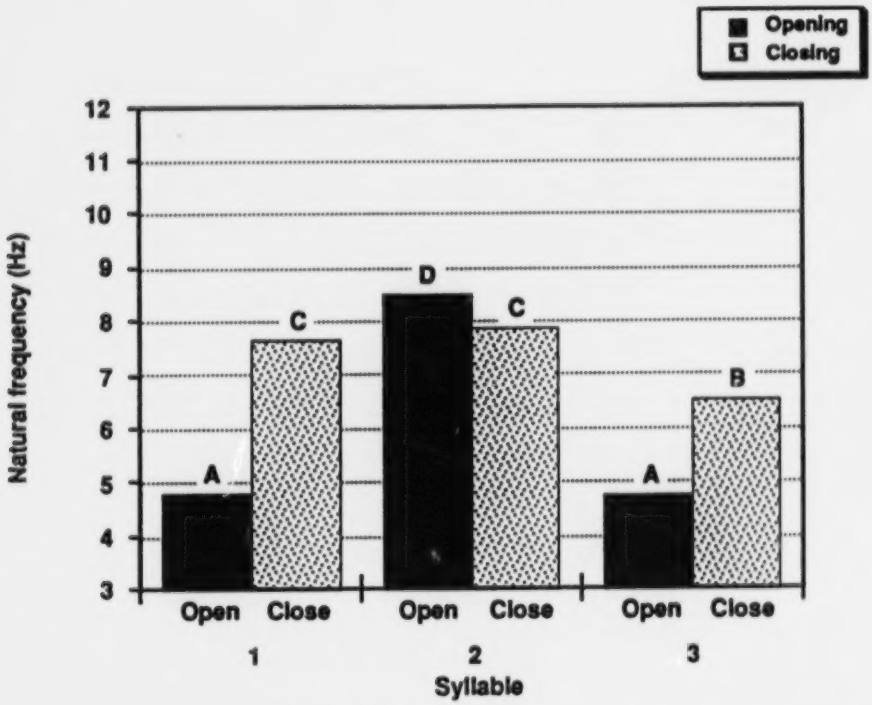


Figure 17. Natural frequency values in the individual gestures of the fits with damping ratio .5. The letters indicate gestures that could be grouped by their significant differences (using Newman-Keuls tests).

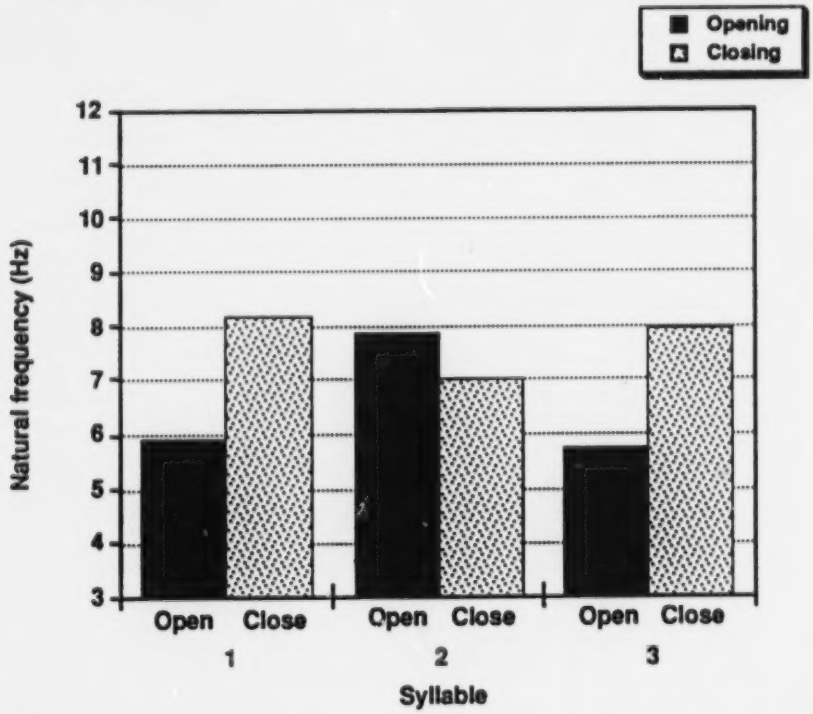


Figure 18. Natural frequency values in the individual gestures of the fits with damping ratio .8. Neither the effect of Syllable nor the interaction of Syllable × Direction was significant everywhere.

For Stress as for the other main effects, the results obtained from the fits at .2 and .5 damping ratios were very similar to the results obtained from the least error fits. Frequency in the stressed condition was significantly lower than in the unstressed in the first two syllables of the utterance in the least error, the .2 and the .5 fits. In the .5 fits, the significance of Stress was limited by Vowel quality, as Stress was significant with both vowels in closing gestures, but in opening gestures only when the full vowel was /a/. In contrast to the lower damping ratios, in the fits at .8 Stress was significant overall. This change seems to reflect a loss in the distinction between the first two syllables of the utterance, where the effect of Stress is robust, and the final syllable, where the difference between stressed and unstressed vanished at the lower damping ratios.

Natural frequency was not significantly affected by the factor of Vowel at any damping ratio. As in the fits with the least error, in the .2 damping ratio fits, only in opening gestures were the frequency values significantly different when analyzed with Peak windows than with CV windows. (Compare Figure 19 to Figure 12.) In contrast, in the .5 fits (Figure 20) the effect of Window Type was significant everywhere; the mean frequency with Peak windows was lower than that for CV windows. In the .8 fits (Figure 21), as was the case with the linguistic factors, Window Type was significant in more limited environments than it was at lower damping ratios. Thus Window Type represented a slight divergence of the .8 fits from the lower damping ratios, although the overall pattern among the values was still much the same.

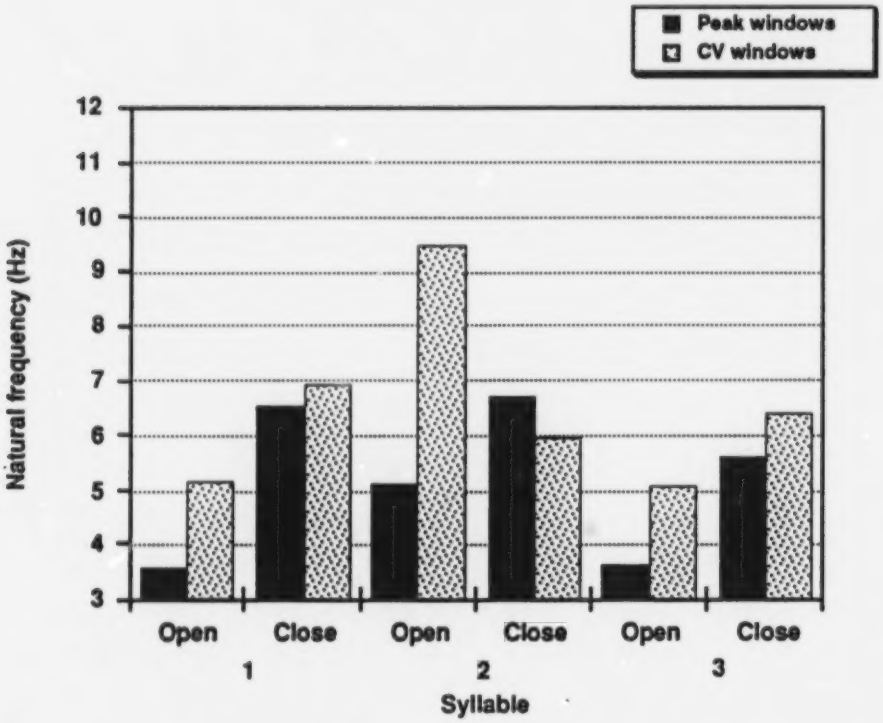


Figure 19. Natural frequency values in the fits with damping ratio .2, showing the effect of Window Type in the individual gestures.

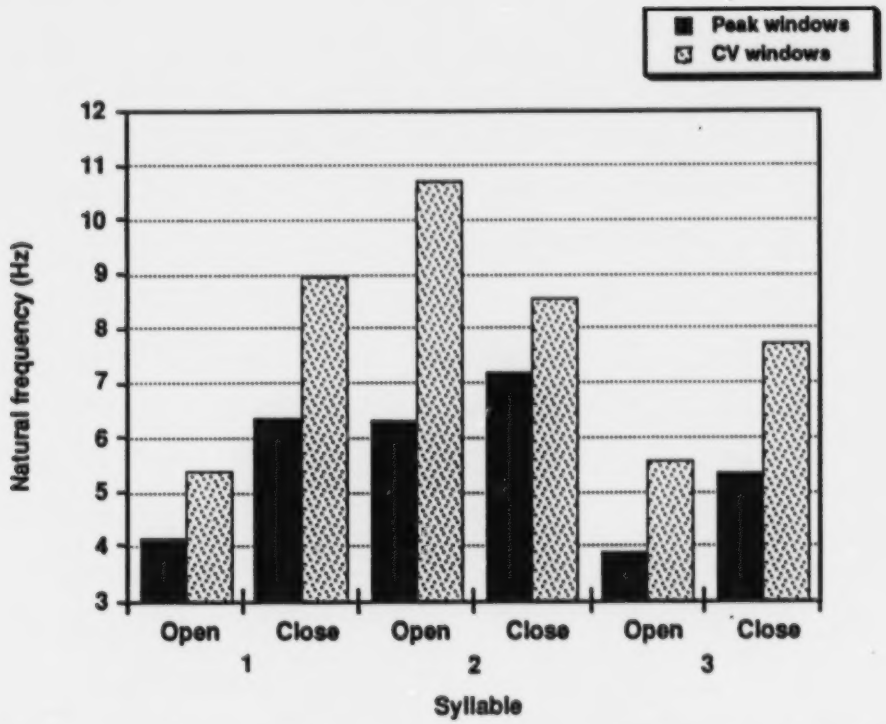


Figure 20. **Natural frequency values in the fits with damping ratio .5, showing the effect of Window Type in the individual gestures.**

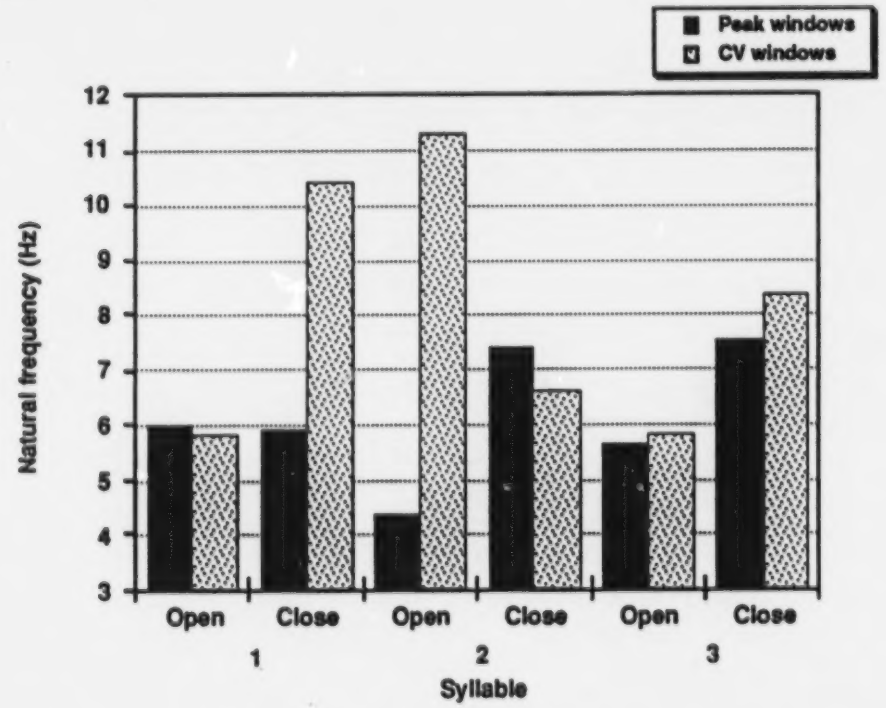


Figure 21. **Natural frequency values in the fits with damping ratio .8, showing the values for the individual gestures analyzed with both Window Types. Neither the effect of Window Type nor the interactions with it were significant everywhere.**

The gesture that was anomalous in the least error fits—the closing gesture in the reduced syllable 2—remained anomalous in the .2 and .8 fits, but was anomalous in the .5 fits only so far as the opening/closing pattern, as can be seen in Figures 12 (least error), 19 (.2), 20 (.5), and 21 (.8). In this gesture, the least error, .2, (and .8, although here significance was severely limited) fits had higher frequencies for Peak analyses than for CV analyses, which was the reverse of the normal Peak/CV relation. The .5 fit differed from the other fits for this gesture in that the analysis with Peak windows returned lower frequencies than the analysis with CV windows. However, the .5 fit was similar to the other fits for this gesture in that this closing gesture had lower frequency than the opening gesture in the same reduced syllable when analyzed with a CV window, unlike the normal pattern in which closing gestures had higher frequency.

3.3.2 Discussion

As the damping ratio of the fit was increased, higher natural frequency values were typically obtained, as expected. That is, when analyzing a given movement trace or curve, the observed frequency is always the same. A given observed frequency can be obtained through various combinations of natural frequency and damping. Decreasing the natural frequency or increasing the damping will both decrease the observed frequency (see equation (4) in the introduction). However, in the different analyses described above, the observed frequency remained the same as the damping ratios changed in the fits. Therefore, as the damping ratio was increased, the natural frequency must perforce increase in order to achieve the same observed frequency. The fact that the program behaved in this way is further confirmation that it is behaving sensibly, rather than a result. However, it also means that the absolute values of any natural frequencies returned from an analysis of this type are unlikely to be an accurate characterization of the data unless the damping ratio used is close to the correct value.

The effects of the linguistic factors on the natural frequency, however, appeared to be robust across the damping ratios, with only a decrease in the stability of the statistical significance when the damping ratio was increased to .8, a value very different from the average damping ratio of the data (.13) as determined by the least error fits. That is, it appears that using an inappropriate damping ratio did not alter the patterns of natural frequency values, but rather made some of them harder to discern. The patterns that held at all the damping ratios included the following:

(1) The frequencies of the three syllables of the utterance were in the same order: syllable 3 the lowest, then 1, then 2 (the reduced syllable).

(2) Opening gestures tended to have lower frequencies than closing gestures.

(3) Stressed gestures tended to have lower frequencies than unstressed gestures, especially in non-final syllables.

(4) Peak windows tended to have lower frequencies than CV windows.

(5) The closing gesture of syllable 1 tended to group with the gestures of the reduced syllable 2 in having high frequencies.

The size of the effects changed somewhat among the different sets of fits, with the .8 fits usually having the smallest and therefore statistically least robust effects. The reduced significance in the statistical results at .8 was due at least in part to the fact that these fits were characterized by much greater variability than at the lower damping ratios. For example, Figure 22 shows the standard deviations for the frequency values of the individual syllables. Although the standard deviation did increase from .2 to .5, a much larger increase occurred between .5 and .8. This increase in variability would have contributed to the more limited significance of the factors.

One example of a major change in significance between the lower damping ratios and .8 was the effect of Syllable. This was significant in the least error fits and at .2 and .5, but not at .8. This change could have resulted from the increase in variability, but also from smaller differences among the frequency values for the various syllables at .8 (shown in Figure 7), possibly due in turn to the relatively high frequencies. Another example of a change in significance involved the change of the Stress effect in the final syllable for the .8 fits. In the .2 fits and the .5 fits, the effect of Stress (lower frequency for stressed relative to unstressed) was, as in the least error fits, limited to the first two syllables of the utterance, presumably because final lengthening reduced this distinction in the last syllable. This contrast between the first two syllables, which had a stress distinction, and the final syllable, which did not, vanished in the .8 fits, in which the effect of Stress was regularized to be the same in all syllables. This change may have resulted from the less accurate fits at .8 not capturing the final lengthening that distinguished the final syllable; this loss of final lengthening might also contribute to the loss of significance of Syllable position in the .8 fits.

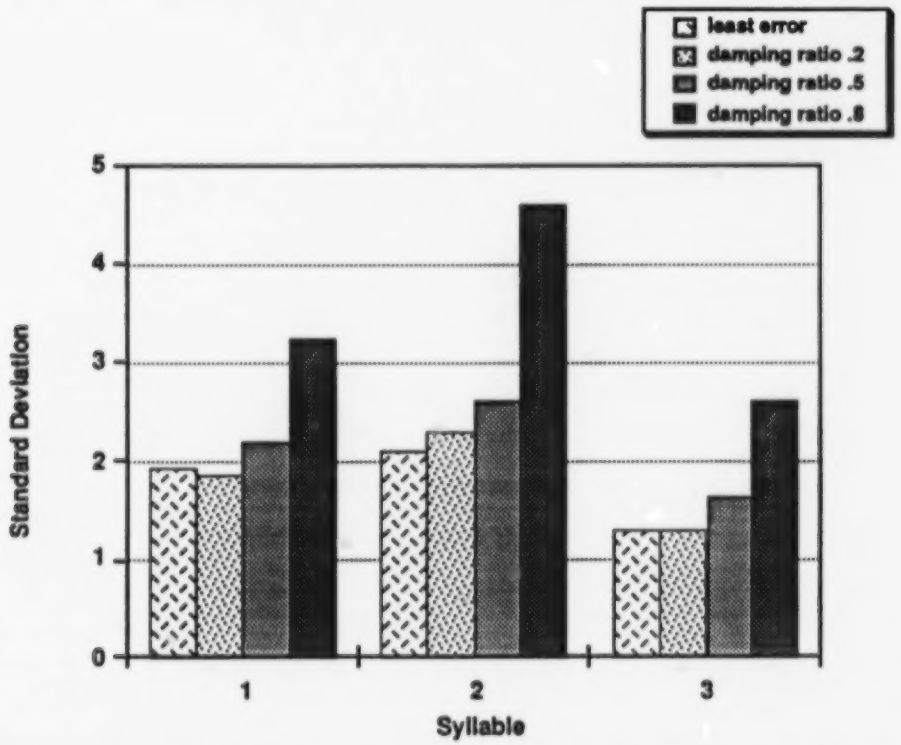


Figure 22. **Standard deviations of the mean values for natural frequency in each of the three syllables of the utterances, in the least error fits and at the three fixed damping ratios.**

4 SUMMARY

We have described a set of procedures for analyzing articulatory movement data, under the assumption that such data can be characterized in terms of a linear second-order dynamical system. In addition, we have presented some results of articulatory analyses performed following these procedures.

Given that a particular curve can be fit reasonably well by a number of different parameter sets, it was necessary to develop procedures for selecting among possible parameters, and to test the reasonableness of the parameters returned by these procedures. Two important components of the procedures developed involved constraining the range of possible values of the parameters and reducing the degrees of freedom. The degrees of freedom were reduced by utilizing some constraining conditions based on selected values of the data, by holding the damping ratio constant and using a least error procedure to select the best fit, and by restricting the combinations of constraining conditions and damping ratios to those with minimal error and/or variance for simulated data. For the simulated data, these procedures returned values for the damping ratio that were within +/- .1 of the correct damping ratio, and values for natural frequency that were within 5% of the correct value, on the average.

When the above procedures were applied to measured articulatory data, in particular to the movements of the lower lip, most of the resulting parameter values appeared to reflect attributes of the data; in a few cases, the returned values appeared to be artifacts of the analysis procedure or the type of analysis window. The increased natural frequencies associated with the CV windows (over Peak windows) and with the higher fixed damping ratios (over lower fixed damping ratios) are the two primary examples of parameter values that are analysis artifacts rather than attributes of the data. In addition, the similarity between frequency and phase angle effects, such that increased frequency tends to co-occur with larger phase angles, may also be an artifact of the analysis procedure.

The effects of the linguistic factors on the natural frequency were remarkably robust across the various damping ratio assumptions (least

error or fixed at .2, .5, .8), and the effects reflected those expected on the basis of other studies. Natural frequency was lower for stressed gestures than for unstressed gestures, even when the damping ratio of the fit (held fixed) was very far away from the least error damping ratio. In general, natural frequency was lower for opening than for closing gestures, even when the damping ratio was highly inappropriate. Among the syllables, the natural frequency was highest for the reduced syllable and lowest for the final syllable across all damping ratios. In addition to these results expected on the basis of other studies, closing gestures tended to show characteristics of the following syllable. This was particularly true for the closing gesture out of the first syllable into the second (reduced) syllable, and also played a role in creating apparent anomalies in the behavior of the closing gesture out of the second (reduced) syllable into the last syllable.

The robustness of the effect of the linguistic factors on natural frequency across damping ratios has several possible implications. From a purely methodological perspective, of course, it is very encouraging. That is, for these data little information was lost by analyzing all the data at a single damping ratio, even when that damping ratio was arbitrarily chosen, since the patterns remained even when the statistical significance weakened. Thus the patterns found by PARFIT were not dependent on the "best" fits being used, but could be generalized to fits at other damping ratios. Such a finding is not only helpful for the purposes of the modelling procedure for which PARFIT was originally developed (which assumes a fixed damping ratio throughout), but also is encouraging in general about the potential reliability of the linguistic effects on natural frequency.

The robustness of the frequency effects across various fixed damping ratios might also partially reflect an inherent lack of effect on the damping ratio by the linguistic factors. That is, if the damping ratio does not inherently vary much, then analyses at fixed damping ratios would not potentially introduce errors due to their fixedness, but only due to their inappropriate numerical values. Certainly for these data, the linguistic factors did not affect the damping ratio much: the damping ratio for the least error fits rarely differed systematically from its average value of 0.13. The primary significant increase in damping was found in CV windows for closing gestures, and could presumably be attributed to compression of the lips during the bilabial closures.

The fact that significant damping ratio differences were detected primarily in CV windows highlights the fact that the choice of analysis window can significantly affect the results obtained. Since virtually all analyses of articulatory movement rely on measuring particular portions of movement trajectories, any difference that arises solely from the method of partitioning the trajectory has important consequences. In general, the effects of the linguistic factors on natural frequency were the same for the two window types (which also serves as a further indication of the robustness of the effects of the linguistic factors on natural frequency). However, one gesture, the closing gesture out of the reduced second syllable into a following non-reduced syllable, showed very different results depending on the type of window used. That is, the difference between the portion of movement captured in a CV window and in a Peak window resulted in a significantly different estimate of natural frequency for the closing gesture in a reduced syllable, and in only the CV windows being sensitive to the increased damping ratio of closing gestures in general.

Further work is needed to investigate the effect of different window types. Since only analyses using windows that contain an entire extremum are apparently able to pick up damping ratio differences, it might be useful to compare windows in which the extrema are at the beginning to those in which the extrema are at the end (the CV windows in the current analyses). Further analysis of other articulators is also needed to determine whether the results shown here are general characterizations for all articulators. We would expect the frequency patterns (although not necessarily the absolute values) to be robust across articulators, but different articulators might well show different characteristic damping ratio behaviors. The procedures developed for the use of PARFIT appear to be reliable and accurate enough to warrant these further studies.

REFERENCES

Browman, C. P., & Goldstein, L. (1985). Dynamic modeling of phonetic structure. In V. A. Fromkin (Ed.), *Phonetic linguistics* (pp. 35-53). New York: Academic Press.

Browman, C. P., & Goldstein, L. (1990). Tiers in Articulatory Phonology with some implications for casual speech.In J. Kingston & M. Beckman (Eds.), *Papers in laboratory phonology I*. Cambridge: Cambridge University Press.

Browman, C. P., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology, 6*, 201-251.

Browman, C. P., Goldstein, L., Saltzman, E., & Smith, C.(1986). GEST: A computational model for speech production using

dynamically defined articulatory gestures. *Journal of the Acoustical Society of America, 80,* S97.

Cooke, J. D. (1980). The organization of simple, skilled movements. In G. E. Stelmach & J. Requin (Eds.), *Tutorials in motor behavior.* Amsterdam: North Holland.

Fowler, C. A., Rubin, P., Remez, R. E., & Turvey, M. T.(1980). Implications for speech production of a general theory of action. In B. Butterworth (Ed.), *Language production.* New York: Academic Press.

Kelso, J. A. S., Saltzman, E. L., & Tuller, B. (1986). The dynamical perspective in speech production: Data and theory. *Journal of Phonetics, 14,* 29-59.

Kelso, J. A. S., V.-Bateson, E., Saltzman, E., & Kay, B.(1985). A qualitative dynamic analysis of reiterant speech production: Phase portraits, kinematics, and dynamic modeling. *Journal of the Acoustical Society of America, 77,* 266-280.

Kent, R. D., & Netsell, R. (1971). Effects of stress contrasts on certain articulatory parameters. *Phonetica, 24,* 23-44.

Kozhevnikov, V. A., & Chistovich, L. A. (1966). *Rech, Artikulyatsiya, i vospriyatiye, [Speech: Articulation and perception]* (originally published 1965). Washington, DC: Joint Publications Res. Service.

McGowan, R., Smith, C., Browman, C., & Kay, B. (1988). Extracting dynamic parameters from articulatory movement. *Journal of the Acoustical Society of America, 83,* S113.

McGowan, R., Smith, C., Browman, C., & Kay, B. (1990). Methods for least-squares parameter identification for articulatory movement and the program PARFIT. *Haskins Laboratories Status Report on Speech Research, SR-101/102,* 220-230.

Munhall, K. G., Ostry, D. J., & Parush, A. (1985). Characteristics of velocity profiles of speech movements. *Journal of Experimental Psychology: Human Perception and Performance, 11,* 457-474.

Oller, D. K. (1973). The effects of position in utterance on speech segment duration. *Journal of the Acoustical Society of America, 54,* 1235-1246.

Ostry, D. J., Keller, E., & Parush, A. (1983). Similarities in the control of speech articulators and the limbs: Kinematics of tongue dorsum movement in speech. *Journal of Experimental Psychology: Human Perception and Performance, 9,* 622-636.

Ostry, D. J., & Munhall, K. (1985). Control of rate and duration of speech movements. *Journal of the Acoustical Society of America, 77,* 640-648.

Saltzman, E., Goldstein, L., Browman, C., & Rubin, P. (1988). Modeling speech production using dynamic gestural structures. *Journal of the Acoustical Society of America, 84,* S146.

Saltzman, E., & Munhall, K. G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology, 1,* 333-382.

Sonoda, Y., & Kiritani, S. (1976). Analysis of tongue point movements by a linear second-order system model. *Annual Bulletin RILP, 10,* 29-35.

Tuller, B., Harris, K., & Kelso, J. A. S. (1982). Stress and rate: Differential transformations of articulation. *Journal of the Acoustical Society of America, 71,* 1534-1543.

Vatikiotis-Bateson, E. (1988). *Linguistic structure and dynamic parameters: A kinematic study of reiterant speech production in three languages.* Bloomington: Indiana University Linguistics Club.

Zawidzka, P. (1981). Tongue-apex activities during alveolar stops. *Phonetica, 38,* 227-235.

FOOTNOTES

†Also Department of Linguistics, Yale University.

††Department of Psychology, Brown University.

Haskins Laboratories Status Report on Speech Research
1991, SR-105/106, 141-152

Phonological Underspecification and Speech Motor Organization*

Suzanne E. Boyce,† Rena A. Krakow,†† and Fredericka Bell-Berti†††

1 INTRODUCTION

Over the last few years, much work in phonology has been devoted to exploring the way features are specified for segments; in particular, to what extent feature specification may be underlyingly present and/or acquired by rule or default in the course of a derivation. While a number of proposals have been made attributing various degrees of underspecification to abstract levels of the phonology (Archangeli, 1988; Kiparsky, 1985; Steriade, 1987), it has been generally assumed that where phonetic implementation comes into play, i.e., at the end of the derivation, segments are exhaustively specified.

This view stands in contrast to that adopted in much of the literature on speech motor control where the necessity of accounting for coarticulation across multisegmental spans has led researchers to assume that the input to the motor plan leaves a good deal of phonetic detail unspecified. Motor implementation in these models is assumed to proceed by direct translation of specified features into articulatory/acoustic targets, leaving the position of articulators during an unspecified segment open to influences from the surrounding context. Thus, coarticulation is viewed as assimilation of specified features from surrounding segments onto an unspecified target. As a practical matter, researchers in the field have assigned feature specification for this purpose from: (1) surface phonological contrast and (2) required articulatory/aerodynamic configurations. If neither source mandates specification, the segment is assumed to be unspecified for that feature and thus to have no independent target.

For example, [+nasal] and [-nasal] are feature values that characterize English stops at the surface level and therefore these consonants are taken to require a relatively low or relatively high velum position, respectively. At the same time, because English vowels lack a surface contrast in nasality and because nasalization of vowels is articulatorily possible (while a nasal /s/ is not), coarticulation researchers have assumed that English vowels are unspecified for nasality (e.g., Kent, Carney, & Severeid, 1974; Moll & Daniloff, 1971). Note that, in this view, English vowels and stop consonants have the same [nasal] specification status underlyingly as at the surface. The case is different for a segment such as /s/. On the one hand, it is presumed to have the surface feature value [-nasal] because a high velum is required to produce the necessary aerodynamic conditions for high intensity frication. On the other hand, English fricatives do not contrast with respect to nasalization so that specification of [nasal] for /s/ may be lacking at more abstract levels of the derivation.

In English as in other languages, some amount of nasalization is generally present on vowels preceding nasal consonants (Clumeck, 1976). In analogy to phonological analyses of assimilation processes such as vowel harmony, many studies of this phenomenon (Hammarberg, 1976; Kent et al., 1974; Moll & Daniloff, 1971) have treated the presence of even minimal acoustic or articulatory

We want to thank Carol Fowler, Marie Huffman, Michel Jackson, Ellen Kaisse, Robert Ladd, Ignatius Mattingly and Sharon Manuel for providing extremely helpful comments on an earlier draft of this paper. Most importantly, we thank Pat Keating for sharing her ideas about the possible relation between underspecification in phonology and phonetics. We see our work as another step in exploring this link and we have benefitted greatly from her comments. We also thank Pat for allowing us to use her spectrograms of the Russian words discussed in her paper and in ours. This work was supported by NIH Grant DC-00121 to Haskins Laboratories and by NIH grant NSO-7040-15 to MIT.

indicators of nasalization as showing the spread of the feature specification [+nasal] into an unspecified domain. It was assumed that the intermediate level of implementation typical of the data came from an inability of the articulators to achieve opposite target configurations instantaneously; that is, it was assumed that the vowels acquired a full [+nasal] target but could not fully implement it, rather than that the vowels had independent but intermediate targets (see Kent et al., 1974 for further discussion).

In a seminal paper, Keating (1988b) examined theories of underspecification as they exist both in the motor control literature and in the phonological literature, in an attempt to reconcile the two. She accepted the motor control notion that there are phonetically unspecified segments, and that these segments have no inherent targets; but argued that, rather than indicating the presence of feature spread, the presence of intermediate levels of articulatory/acoustic implementation (e.g., slight lip protrusion for rounding in the context of a [+round] segment, or the presence of a weak nasal formant in the context of a [+nasal] segment) could be taken to indicate persistence of underspecification into the motor planning level. Further, she suggested that if a segment normally analyzed as unspecified for a particular feature showed an apparent target, i.e., showed apparent full implementation for that feature, a phonetic rule must have applied to supply that target. She proposed that segments are not exhaustively specified at the end of the derivation, and most radically, and interestingly, that phonetic data can be used to make inferences about the lack of specification at higher levels of the derivation.

In this paper, we examine some of Keating's arguments and introduce data of our own indicating that certain of her conclusions may be premature and/or insufficiently detailed. In particular, we attempt to show that, although Keating's basic insight remains viable, much of her argument suffers from the nature of her assumptions about the phonetic implementation of targets, and that a greater attention to phonetic detail, and in particular to the variable of timing, is required in order to eliminate other interpretations of phonetic data. In opposition to Keating's point of view, we present evidence to show that segments which lack specification by contrast crit ria or by aerodynamic/articulatory criteria nevertheless exhibit characteristic articulatory positions associated with those features. Our data will focus on the features [round] and [nasal].

Figure 1(a and b) illustrates, in schematic form, what we take to be the essentials of Keating's model for articulation. What is sketched is the time course of velum movement, which is considered to be a fairly direct index of the feature [nasal] (Keating, 1988a).[1] Here, feature specification translates into targets for articulators, and the motor program moves between targets by simple linear interpolation.[2] When two segments with opposite specifications occur in sequence, the transition between their opposite (and relatively extreme) articulator positions is necessarily speedy and steeply pitched (Figure 1a).

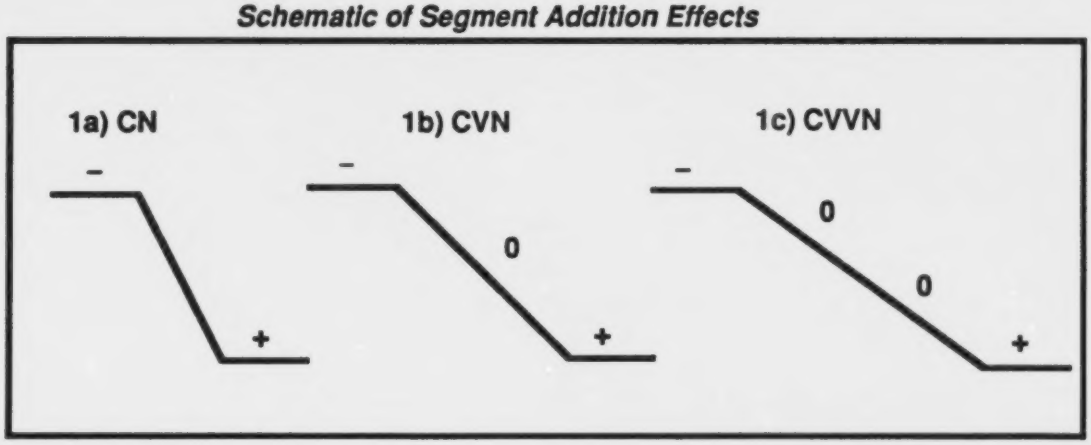


Figure 1. **Schematic diagrams showing transitions between fully specified segments separated by (A) no segments; (B) one unspecified segment; (C) two unspecified segments. In this example, [+] and [-] values refer to the feature specifications for [nasal] and the trajectory corresponds to vertical velum displacement with high velum positions at the top and low velum positions at the bottom.**

Unspecified segments, while introducing an additional timing unit, have no targets and therefore "when phonetic rules build trajectories between segments, an unspecified segment will contribute nothing of its own to the trajectory" (Keating, 1988b: 281). Keating thus predicts that, in these cases, interpolation between [+] and [-] segments will take longer and be less steep than when the two specified segments are immediately adjacent. In essence, the transition between immediately adjacent segments is 'stretched' when they are separated by one or more unspecified segments (Figure 1b and 1c).

Patterns resembling the 'stretched' transitions predicted by Keating's model are easily found in acoustic and articulatory data of all kinds.[3] Keating argues that apparent examples of stretched transitions provide evidence for phonological underspecification continuing to be present at the end of the phonological derivation, i.e., the input to the motor plan. In essence, her argument is as follows: if a segment shows what looks like a target, then it is safe to assume that it enters the motor program with categorical specification (i.e., either [+] or [-]). If, on the other hand, the data show smooth interpolation-type transitions between specified segments, then it is safe to infer that the intervening segments have no targets. If they have no targets, then they are unspecified at the level where phonological representation is translated into a motor program. In what follows, we will refer to Keating's model as the 'Target Equals Specification' hypothesis, or TES.

If we accept Keating's reasoning, there are several tests that can be done to determine the validity of this model. First, because the TES theory crucially depends on unspecified segments having no target for the feature of interest, it is necessary to establish in any particular case that there is no evidence of such a target. (It is always possible, for instance, that an apparently smooth transition is passing through a target.) There are a number of reports in the literature indicating that for many segments, although contrast arguments and articulatory/aerodynamic arguments for specification do not exist, characteristic articulatory targets may be observed across contexts. It is well known, for example, that the English consonants /r/, /s/ and /ʃ/ often show rounding (Brown, 1981). Similarly, the vowels of English are well known to show positions of velum height that are midway between the very high and very low positions of oral and nasal consonants, respectively, and that differ as a function of vowel height (e.g., Bell-Berti, Baer, Harris, & Niimi, 1979; Ohala, 1971). If such findings are typical, then sequences of the type Keating discusses may show evidence of independent targets. One way to test this is by comparing the segment in minimally contrastive contexts, i.e., varying the specification of flanking segments one segment at a time.

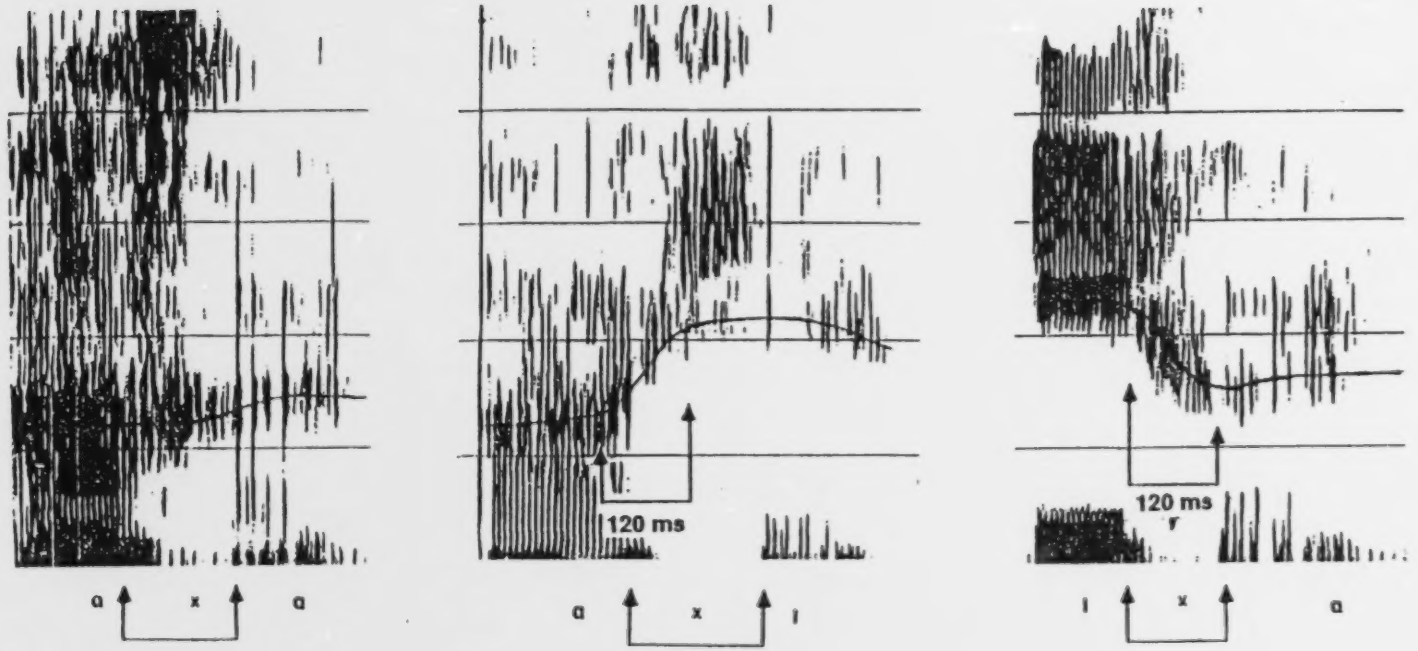
2.1 Minimal contrasts

An example from Keating (1988b) may make this procedure clearer. Although other fricatives in Russian show a surface contrast for palatalization, /x/ does not. Thus, it should be unspecified for the feature [back]. Figure 2 reproduces spectrograms from Keating (1988b) showing /x/ in the context of two back vowels, /ɑxɑ/, plus the complementary back and front contexts, /ɑxi/ and /ixɑ/.

Noting that the second formant (which reflects tongue positioning in the front-back dimension) appears to be in continuous movement through the occluded portion of /ixɑ/ (from the high position for /i/ to the low position for /ɑ/), Keating argues that /x/ cannot have a target and thus must be underspecified at the motor implementation level. This case contrasts with that for /ɑxi/, where the second formant reaches an /i/-like steady state midway through the occluded portion of /x/. Keating argues that, in the case of /ɑxi/, a context-sensitive phonetic rule has applied, spreading the [-back] specification of /i/ into /x/.

Crucially, the TES (Target Equals Specification) hypothesis requires that underspecified segments contribute nothing to the articulatory plan (that is, with regard to the feature of interest). In order to establish that, in /ɑxi/, the high second formant seen during the occluded portion of the signal is due to a [-back] specification copied from the /i/, it is necessary to show that /x/ does not normally show a high second formant in other, non-fronted contexts. On the other hand, if /x/ is unspecified, the model predicts that the interpolation between two [+back] vowels will be a straight line. Keating's data for /ɑxɑ/, where the flanking vowels have [+back] specifications, in fact show that /x/ in this context has a low second formant similar to the formants for the two /ɑ/'s. This is consistent with her claim that the unspecified segment /x/ has no target.

One problem with the conclusion that a feature copying rule has applied, however, is that the slopes of the transitions between /ɑ/ and /i/ in /ɑxi/ and between /i/ and /ɑ/ in /ixɑ/ are too similar to correspond to the predictions of the TES model.



154 *Figure 2.* **Spectrograms adapted from Keating (1988b) showing formant tracks for Russian /ɑxɑ/, /ɑxi/, /ixɑ/. The time course of the second formant, which is considered a reasonable index of tongue constriction along the front/back dimension, is traced by a solid line. The figures are modified from Keating's originals by inclusion of brackets indicating (1) beginning and end of the occlusion period for the /x/, and (2) beginning and end of the second formant transitions between vowel steady states.** 155

In Keating's analysis, the /x/ of /ixɑ/ is unspecified while the /x/ of /ɑxi/ is specified. Thus, formant movement in the two cases should correspond to the schemata of Figures 1b and 1a, respectively, in that the transition for /ixɑ/ should be longer and less steep than that for /ɑxi/. Although for /ixɑ/ the second formant transition between /ɑ/ and /i/ takes place largely during the occluded portion of /ixɑ/, and for /ɑxi/ is divided between the final portion of the /ɑ/ and the occluded portion, the actual durations of the transition are almost identical. This is hard to reconcile with the TES model, where the presence of an unspecified segment, by leaving the tongue with no target articulation, necessarily causes it to move more slowly between specified segments. We suggest an alternative explanation, i.e., that the timing of movement for the tongue (resulting in movement of the second formant) is similar in the two cases, but merely starts earlier in the first vowel for /ɑxi/.[4] This may have the effect of fronting /x/ (and giving the listener the perception that a phonetic rule fronting /x/ was intended by the speaker), but is not an example of feature copying in the sense in which Keating uses it.

As far as they go, these data concerning Russian /x/ suggest that /x/ is indeed unspecified with regard to [back]. It is still possible, however, that a target for /x/ exists, but is not visible in the time course of the second formant because the tongue 'en route' to /ɑ/ or /i/ does not have time to show this independent target. A good way to test for such a target is to insert additional 'unspecified' segments to see whether the transition between the flanking specified segments becomes lengthened and less steep. (Keating makes no mention of her expectations for instances in which the number/duration of unspecified segments is increased; in fact, the contribution of time to the model is not addressed.) Thus, test sequences should be constructed so as to maximize the opportunity for articulatory/acoustic behavior to show itself. This is schematized in the series a-b-c of Figure 1. Note that while the slope of the transition should change, the transition itself should remain smooth. Another, equally valid, test is to decrease speaking rate (i.e., slow down). This likewise, by increasing the time gap between specified segments with opposing values, ought to lead to lengthened and more gradual, but smooth, transitions (see Figures 3a and b).

In what follows, we will test these notions using data from two articulators, the lip and the velum, as related to rounding and nasalization, respectively. We will show, by adding time to the trajectories via segment addition and speech rate manipulation, that evidence for underspecification of the type Keating discusses may be more apparent than real. Further, we will advance a very different argument about what appear as smooth trajectories through unspecified segments. Our claim is that, at least for these data, independent targets for so-called unspecified segments exist although temporal constraints may prevent them from being visible in the acoustic or articulatory signal.

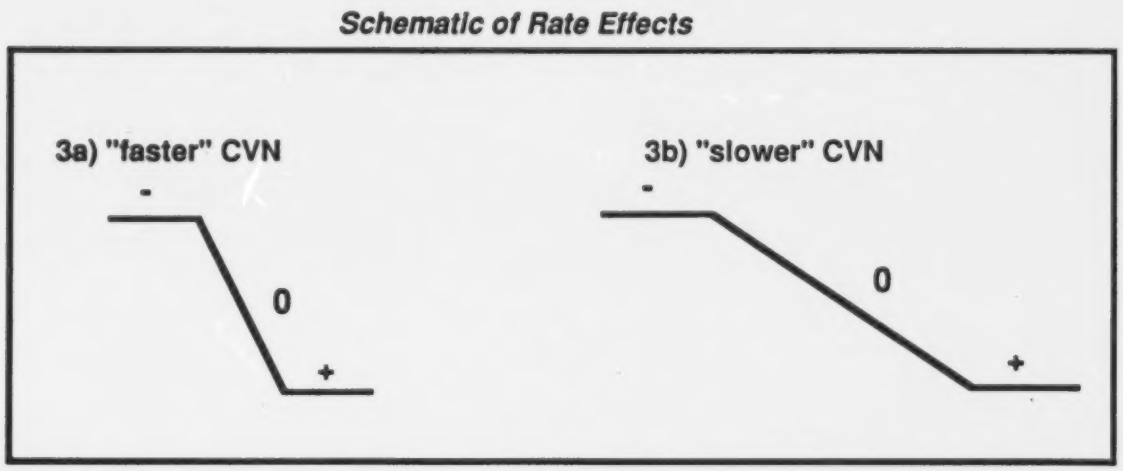


Figure 3. **Schematic diagrams showing transitions between fully specified segments separated by one unspecified segment at (A) faster; and (B) slower rates. [+] and [-] values refer to the feature specifications for [nasal] and the trajectory corresponds to vertical velum displacement with high velum positions at the top and low velum positions at the bottom.**

2. 2 Adding time by adding segments

We begin by examining data on lip protrusion for rounded vowels in American English from a larger study reported in Boyce (1988, 1990). In this study, an optoelectronic tracking system (modified Selspot) was used to track the horizontal movements of a light-emitting diode (LED) attached to the vermilion border of the lower lip. Four native speakers of American English produced fifteen tokens, at a normal rate, for each of a set of nonsense words with various combinations of /k/, /t/, /l/, and the vowels /i/ and /u/. Each was embedded in the carrier phrase "It's a ____ again." Words with /i/ and /u/ vowels (/iC$_n$u/) were chosen to illustrate the case where a segment with [-round] specification (/i/), and a segment with [+round] specification (/u/), are separated by a segment or segments with no specification for [round]. (Again, specification here refers to surface specification.) Words with two /i/ vowels (/iC$_n$i/) were chosen to illustrate the minimal contrast condition whereby segments with no predicted specification for [round] are nevertheless examined for evidence of rounding. Words with two and three intervocalic consonants were chosen to test the prediction, derived from the TES hypothesis, that a longer time interval between [-round] and [+round] segments would result in a longer, more gradual transition. The three panels of Figure 4 show movement traces for, respectively, the nonsense words pairs /kituk-kitik, kiktuk-kiktik, kiktluk-kiktlik/.[5] The traces shown belong to single tokens typical of one speaker's production.

We will examine the results of /iC$_n$u/ words first. As any model of motor control would predict, each of the words in Figure 4 shows clear, relatively extreme, protrusion peaks associated with the [+] valued /u/, and local valleys with the [-] valued /i/. For the word /kituk/, we observe a smooth transition from the [-round] /i/ to the [+round] /u/ through the /t/. This is consistent with Keating's conjecture of linear interpolation between [-] and [+] valued segments through segments with unspecified features. The pattern seen for /kiktuk/ and /kitkluk/, in which the trajectory has acquired local peaks preceding the larger peaks associated with the /u/, is not predicted.

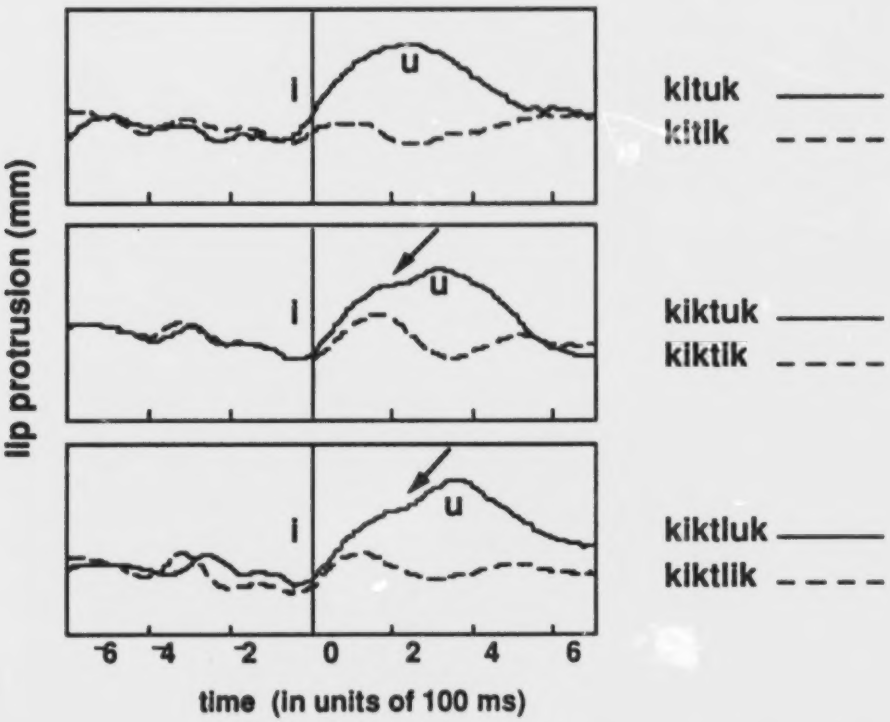


Figure 4. Lower lip protrusion traces for Subject 1's productions of /kituk-kitik/, /kiktuk-kiktik/, /kiktluk-kiktlik/, aligned at the beginning of the intervocalic consonant closure. Upwards movement indicates increased protrusion. Arrows are used to indicate the location in the protrusion trace where a local peak emerges before the protrusion peak for the /u/.

If we look at the /iC$_n$i/ words, however, we see a possible explanation. Movement traces for /kitik/, /kiktik/ and /kiktlik/ all show local peaks in protrusion (flanked by local valleys due to retraction for /i/ vowels). For the pairs /kiktuk/-/kiktik/ and /kiktluk/-/kiktlik/, the peaks in the traces for /iC$_n$i/ words correspond in time to local peaks seen in /iC$_n$u/ words. For /kituk/-/kitik/, the local peak in /kitik/ has no obvious correlate in the smooth movement trace of /kituk/.[6,7] The most perspicuous explanation of these facts is that the intervocalic consonant(s) have independent targets for protrusion. These targets, which are best seen in the traces for /kitik/, /kiktik/ and /kiktlik/, are also evident in the local peaks seen in /kiktuk/ and /kiktluk/. These targets are not visible in the shortest /iC$_n$u/ word (/kituk/) because there is insufficient time for the intervocalic consonant to show a target independent of the trajectory for the rounded vowel. (Note that there remains a slight difference in target position for the consonant(s) in the /iC$_n$u/ and /iC$_n$i/ words.)

A similar argument can be made in the case of what appears to be feature copying. Figure 5 shows characteristic tokens of the word pairs /kituk-kitik/, /kiktuk-kiktik/, and /kiktluk-kiktlik/ for a second speaker. Here the articulatory pattern shows a peak equal to that for the rounded vowel during the consonant(s) for ll three /i-u/ words. Thus, for this speaker, it appears that rounding has occurred on the supposedly unspecified segments. By analogy with Keating's /αxi/ example (above), it might seem that the consonants have copied a rounding feature from the following /u/. On this account, Speaker 2 has a phonetic rule of feature copying for rounding, and Speaker 1 does not.

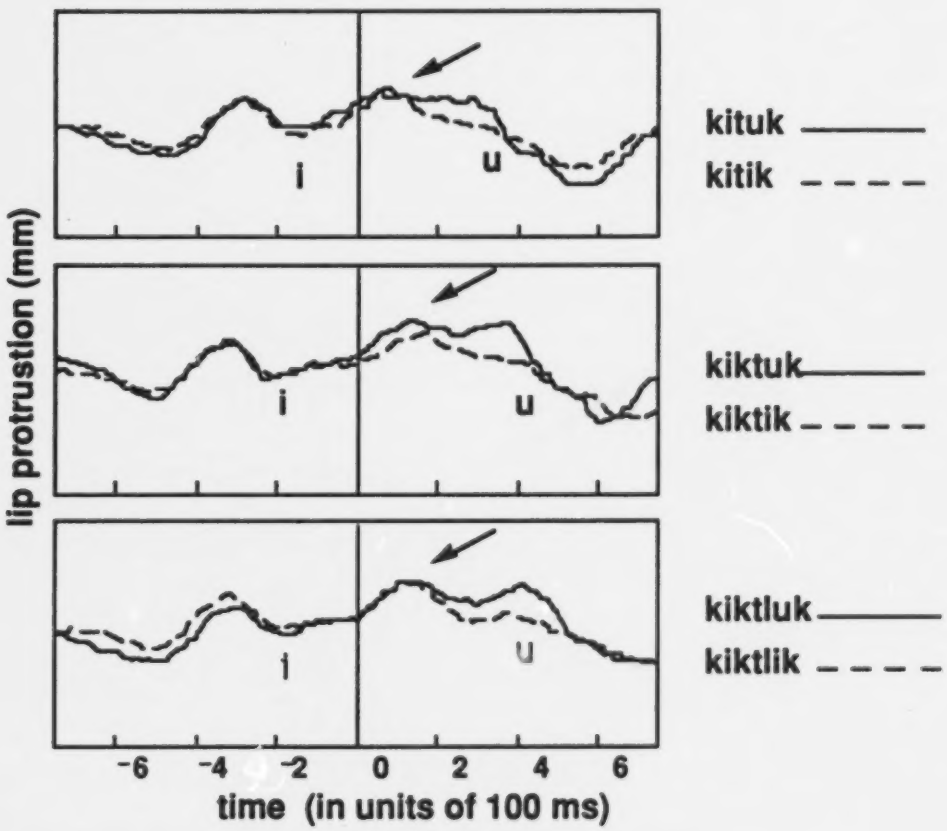


Figure 5. Lower lip protrusion traces for Subject 2's productions of /kituk-kitik/, /kiktuk-kiktik/, /kiktluk-kiktlik/, aligned at the beginning of the intervocalic consonant closure. Upwards movement indicates increased protrusion. Arrows are used to indicate the location in the protrusion trace where a local peak emerges before the protrusion peak for the /u/.

However, both speakers' /iC$_n$i/ words show coincident peaks during the consonant interval. Thus, the more likely explanation is that for both speakers, some or all of the unspecified consonants have targets. These targets are different for the two speakers, such that Speaker 2's consonant(s) are relatively more protruded when compared with their rounded vowels. It is worthwhile noting that although the TES model (and similar models) can account for the behavior of Speaker 2, by postulating a default rule assigning rounding to /t/, models of this type have no obvious way of accounting for less than full rounding such as that exhibited by Speaker 1.[8]

Similar problems arise for the TES model in considering the feature [nasal]. Here we present data from a study of 3 subjects by Bell-Berti and Krakow (1991) that included 12 tokens each of a set of minimally contrastive sequences containing some number of vowels (sometimes in combination with /l/) followed by either another oral consonant (/s/) or a nasal consonant (/n/), including /əsəl/, /ənsəl/, /ləsəl/, /lənsəl/, /ə əsəl/, /ə ənsəl/, /ə ləsəl/, /ə lənsəl/, /se^1 əsəl/,/se^1 ənsəl/, /se^1 ləsəl/, /se^1 lənsəl/. Each of these sequences was preceded by an oral consonant, /s/, in the carrier phrase, "It's _____ again." The Velotrace, a mechanical device developed by Horiguchi and Bell-Berti (1987), was used to monitor the vertical movements of the velum with the aid of a modified Selspot System.

Figure 6 shows the characteristic patterns of velum movement for four sequences containing the post-vocalic nasal consonant /n/ produced at a self-selected rapid speech rate. What appear as smooth interpolation trajectories (of the sort described by Keating, 1988b) are clearly seen in these examples. That is, the velum moves smoothly and continuously through a sequence of intervening vowels (with or without an /l/ in the sequence) between the high velum position required for the /s/ of the carrier phrase and the low position required for the /n/. In general, as the string lengthens, the trajectory appears to stretch. From these data, it might be concluded that the smooth movements indicate a lack of specification for the feature [nasal]. (Note that this conclusion, if drawn, must therefore apply to /l/ as well as to the vowels. See also Moll and Daniloff (1971) for data indicating that the behavior of /l/ resembles that of vowels with respect to velum positioning.)

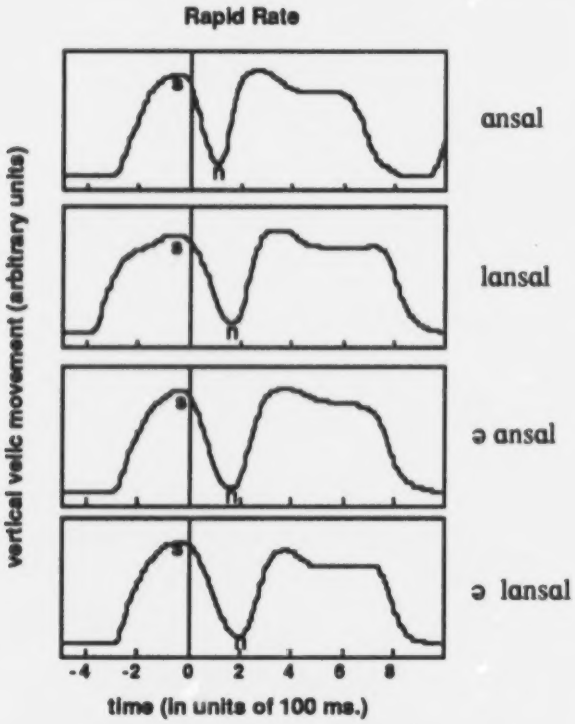


Figure 6. Velum movement traces for /ənsəl/, /lənsəl/, /ə ənsəl/, /ə lənsəl/ produced at a self-selected rapid rate, and aligned at the offset of an immediately preceding /s/ of the carrier phrase in which they were embedded. Target positions associated with the /s/ of the carrier phrase and the /n/ of the sequence are identified. Downwards movement indicates velum lowering (and thus opening of the velopharyngeal port).

2. 3 Adding time by slowing the rate of speech

We would, however, like to approach these movements as we have approached the lip rounding data above: that is, by testing the hypothesis that the shapes of the smooth trajectories represent the combined influences of the sequence of segments during which they are observed. In this case, we claim that the smooth trajectory between /s/ and /n/ is composed of lowering towards specified vowel-, /l/-, and /n/-related velum targets. To support this claim, we compare the utterances of Figure 6, which were produced at a relatively rapid rate, with those in Figure 7 for the same utterances produced at a somewhat slower rate, i.e., the subject's self-selected normal rate. Proceeding from the top to the bottom of Figure 7 we see the effects of adding segments and/or syllables, which result in increasingly clear evidence of an intermediate velum position between the high position of the /s/ and the low position of the /n/. Thus, as we increase the duration of the intervening string between the /s/ and the /n/ (a) by slowing the rate and/or (b) by adding segments/syllables, we begin to see the separate lowering movements that, in faster and shorter sequences remain merged in the movement trace.

These examples provide evidence of a target between the high position for the /s/ and the low position for the nasal consonant. Given these data, the contributions of the individual intervening segments cannot be separated. Bell-Berti and Krakow (1991), however, showed that additional intermediate vocalic targets are observable in the slowest sequences that they examined with multiple vowels. This is consistent with other studies suggesting that there are characteristic positions of the velum for different vowels (Bell-Berti et al., 1979; Henderson, 1984; Kent et al. 1974; Moll, 1962; Ohala, 1971; Ushijima & Sawashima, 1972).

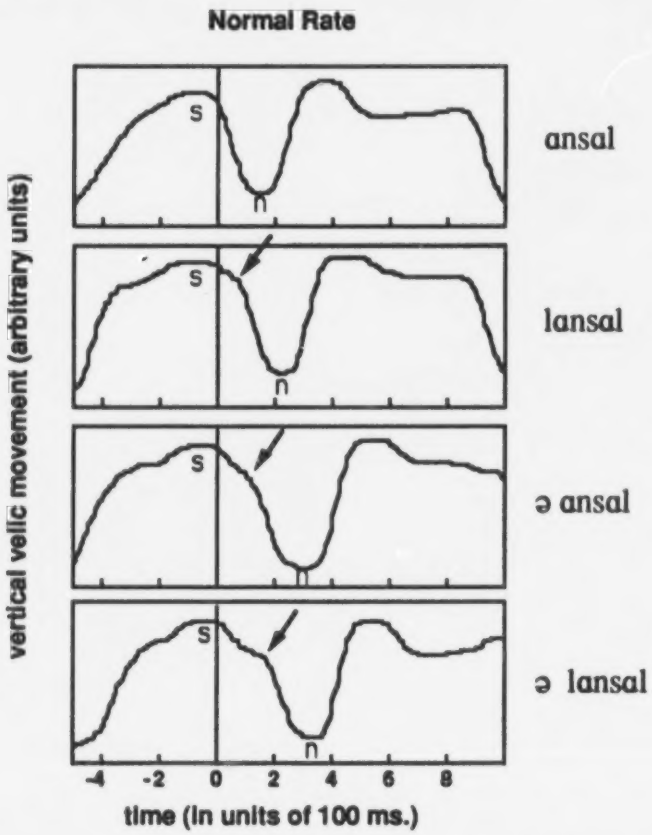


Figure 7. Velum movement traces for /ansal/, /lansal/, /ə ansal/, /ə lansal/ produced at a self-selected normal rate, and aligned at the offset of an immediately preceding /s/ of the carrier phrase in which they were embedded. Target positions associated with the /s/ of the carrier phrase and the /n/ of the sequence are identified. Arrows are used to show the location in the lowering movement at which a separation between two components of that movement may be seen. Downwards movement indicates velum lowering (and thus opening of the velopharyngeal port).

One question that we have not yet answered is whether the intermediate positions of the velum observed following the /s/ are related in some fashion to the upcoming nasal consonant. To test this possibility, we compared minimally contrastive utterances with and without a postvocalic /n/. An example can be seen in Figure 8, where three typical tokens of /ə ləsəl/ and /ə ləsəl/ produced at the self-selected normal rate are paired. The early portion of velum lowering from the high position for the /s/ is much the same across the two contexts, indicating that the intervening string has a specification independent of that for the upcoming oral or nasal consonant. Thus the velum data, like the lip data, indicate that the appearance of smooth trajectories can obscure the presence of specified intervening targets.[9]

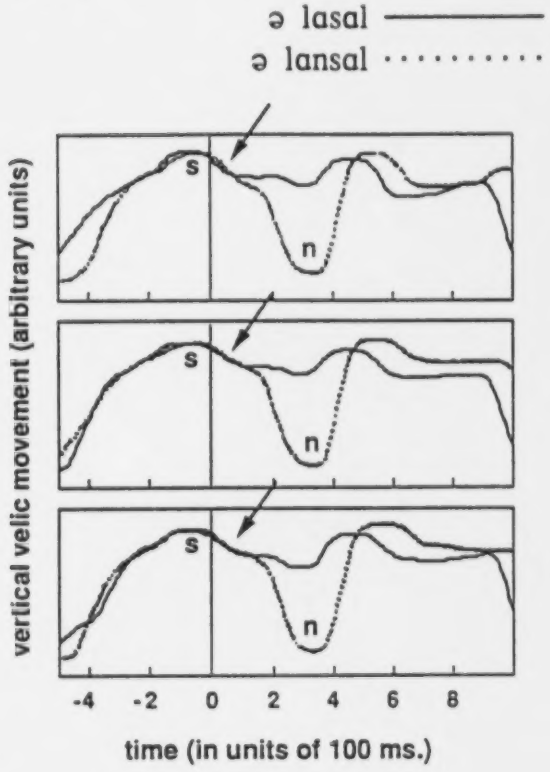


Figure 8. **Velum movement traces for paired tokens of /ə ləsəl/ (solid lines) and /ə lənsəl/ (dotted lines) produced at a self-selected normal rate, and aligned at the offset of an immediately preceding /s/ of the carrier phrase in which they were embedded. Target positions associated with the /s/ of all tokens is shown as is the target for /n/ in those tokens in which it occurs. The similarity between the early portions of the lowering movement in /ə ləsəl/ and /ə lənsəl/ can be seen in the area marked by the arrows. Downwards movement indicates velum lowering (and thus opening of the velopharyngeal port).**

3 CONCLUSIONS

To summarize, the evidence presented here speaks to several issues. Perhaps most importantly, it suggests that segments which lack specification for rounding or nasalization by contrast criteria or by aerodynamic/articulatory criteria nevertheless may exhibit characteristic articulatory positions. However, these positions may be obscured because of temporal constraints. Experimental manipulations that alter or remove these constraints (e.g., comparison between minimal contrasts, adding additional unspecified segments, slowing speaking rate, etc.) are necessary to fully evaluate articulatory behavior that results in a smooth trajectory. Thus, the TES model, in which smooth trajectories are taken as evidence for lack of specification in intervening segments, is not supported. It remains an empirical question, however, whether observed smooth transitions through 'unspecified' segments reflect lack of target specification(s), as appears to be the case for Russian /x/, or a merged and invisible target.

With regard to Keating's original attempt to marry motor and phonological organization, it is not clear how such characteristic articulatory targets for supposedly unspecified segments should be treated. On the one hand, there seems to be a qualitative difference between the type of specification implied by such targets and specification originating at a deeper phonological level. On the other hand, demonstration of a target of any kind is hard to reconcile with the classic notion of underspecification.

One way to interpret the presence of these characteristic articulatory positions is as support for the notion that segments acquire exhaustive specification, by some phonetic evaluation process, just before input to the motor plan. In this view, the notion of phonetic underspecification would have to be abandoned. However, there is another way in which underspecification may influence production. We might think, for instance, that phonologically underspecified features, while associated (in production) with particular articulatory positions for a particular speaker, may be associated with cross-speaker and cross-dialectal variability. The different degree of protrusion evinced by the two English speakers during unspecified consonants (shown in Figures 4 and 5) is a case in point. Similar variability in rounding among speakers has been noted by Brown (1981) and by Gelfer, Bell-Berti, and Harris (1990).

Some of the objections to Keating's (1988b) model described here are met in her "windows" paper (1988a). In that paper, for example, she

proposes limits on velar height for English vowels (in the form of "windows") to a region (vertical dimension) which allows for substantial variability, but nonetheless excludes the extreme low and extreme high positions that are associated with nasal consonants and oral consonants, respectively. For a given speaker, our evidence suggests that there may be more precise articulatory configurations associated with underspecified segments than that paper implies.[10] But perhaps, a "windows" approach may be more easily applied to cross-speaker and cross-dialectal variability in the realization of phonologically unspecified features. Of course this hypothesis will need to be tested empirically as well.

REFERENCES

Archangeli, D. (1988). Aspects of underspecification theory. *Phonology, 5*, 183-207.

Bell-Berti, F., T. Baer, K. Harris, & S. Niimi. (1979). Coarticulatory effects of vowel quality on velar function. *Phonetica*, 36, 187-193.

Bell-Berti, F., & Krakow, R. A. (1991). Anticipatory velar lowering: A coproduction account. *Journal of the Acoustical Society of America, 90*, 112-123.

Boyce, S. E. (1988). *The influence of phonological structure on articulatory organization in Turkish and in English: Vowel harmony and coarticulation.* Doctoral dissertation, Yale University.

Boyce, S. E. (1990). Coarticulatory organization for lip-rounding in Turkish and in English. *Journal of the Acoustical Society of America, 88*, 2584-2595.

Brown, G. (1981). Consonant rounding in British English: The status of phonetic descriptions as historical data. In R. E. Asher & E. J. A. Henderson (Eds.) *Towards a history of phonetics* (pp. 67-76). Edinburgh: Edinburgh University Press.

Clumeck, H. (1976). Patterns of soft palate movement in six languages. *Journal of Phonetics, 4*, 337-351.

Engstrand, O. (1981). Acoustic constraints of invariant output representation? An experimental study of selected articulatory movements and targets. *Report of the Uppsala Univ. Dept. of Linguistics, 7*, 67-94.

Fowler, C. A., Munhall, K., Saltzman, E., & Hawkins, S. (in press). Acoustic and articulatory evidence for consonant-vowel interactions. *Journal of Phonetics.*

Gelfer, C. E., Bell-Berti, F., & Harris, K. S. (1990). Determining the extent of coarticulation: Effects of experimental design. *Journal of the Acoustical Society of America, 86*, 2443-2445.

Hammarberg, R. (1976). The metaphysics of coarticulation. *Journal of Phonetics, 4*, 353-363.

Henderson, J. B. (1984). *Velopharyngeal function in oral and nasal vowels: A cross-language study.* Doctoral dissertation, University of Connecticut.

Horiguchi, S., & F. Bell-Berti. (1987). The Velotrace: A device for monitoring velar position. *Cleft Palate Journal, 24*, 104-111.

Keating, P. A. (1988a). The window model of coarticulation: Articulatory evidence. *UCLA Working Papers in Phonetics, 69*, 3-29.

Keating, P. A. (1988b). Underspecification in phonetics. *Phonology, 5*, 275-292.

Kent, R. D., Carney, P. J., & Severeid, L. R. (1974). Velar movement and timing: Evaluation of a model for binary control. *Journal of Speech and Hearing Research, 17*, 470-488.

Kiparsky, P. (1985). Some consequences of lexical phonology. *Phonology 2*, 85-138.

Klatt, D. H. (1975). Vowel lengthening is syntactically determined in connected discourse. *Journal of Phonetics, 3*, 129-140.

Lubker, J., & Gay, T. (1982). Anticipatory labial coarticulation: Experimental, biological, and linguistic variables. *Journal of the Acoustical Society of America, 71*, 437-447.

Moll, K. L. (1962). Velopharyngeal closure on vowels. *Journal of Speech and Hearing Research, 5*, 30-37.

Moll, K. L., & Daniloff, R. G. (1971). Investigation of the timing of velar movements during speech. *Journal of the Acoustical Society of America, 50*, 678-684.

Ohala, J. J. (1971). Monitoring soft palate movements in speech. *Project in Linguistic Analysis, 13*, JO1-JO15.

Perkell, J. S. (1986). Coarticulation strategies: Preliminary implications of a detailed analysis of lower lip protrusion movements. *Speech Communication 5*, 47-68.

Steriade, D. (1987). Redundant values. *Chicago Linguistic Society, 23*, 339-362.

Ushijima, T., & Sawashima, M. (1972). Fiberscopic examination of velar movements during speech. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics, 6*, 25-38.

FOOTNOTES

*To appear in *Phonology*.

†MIT, Cambridge, MA

††Also Temple University, Philadelphia, PA.

†††Also St. John's University, Jamaica, NY.

[1]Degree of nasality is ultimately a perceptual quality whose most direct articulatory index is the size of velopharyngeal port opening. The vertical position of the velum reflects velopharyngeal port opening (see Horiguchi & Bell-Berti, 1987).

[2]Keating uses linear interpolation in her exposition. She suggests that other types of interpolation may be possible, but their precise manifestation is not explored.

[3]Although her 1988b paper presents acoustic data only, it is clear from this and other writings (e.g., Keating, 1988a) that her hypothesis is a production hypothesis and her conclusions are meant to apply to articulatory data as well.

[4]It's noticeable that the duration of /ɑ/ is longer than that of /i/ in both /əxi/ and /ixɑ/. This is part of a well-known pattern in which higher vowels have shorter intrinsic durations than lower vowels (Klatt, 1975). Also, several researchers have noted a reciprocal temporal relation between adjacent consonants and vowels such that a given consonant is likely to be shorter when its tautosyllabic vowel is longer and vice versa. We clearly see this sort of relation in the longer /x/ preceding /i/ vs. /ɑ/ (Fowler, Munhall, Saltzman, & Hawkins, in press). The time course of tongue movement could, in fact, be the same in both cases, but intrinsic duration differences and the compensatory behavior described might cause the tongue to appear to be moving during the /ɑ/ of /əxi/ but during the /x/ of /ixɑ/. (We are grateful to Sharon Manuel for bringing the issue of timing in these data to our notice.)

[5]Speakers were provided with the real word model "tactless" for the /ktl/ sequence in these items.

[6]The extent to which small variations in movement, as seen in the retraction of the lips for /i/, the intervocalic consonant protrusion movements for /iC$_n$i/ words, and the early peak in protrusion seen for /iC$_n$u/ words, affects the actual perception of features such as [round] cannot be assessed without data from other lip dimensions and/or perceptual data. It should be noted, however, that the range of movement seen for this subject, e.g., 10-12 mm for protrusion related to /u/, and 3-4 mm for intervocalic protrusion in /kiktik/ and /kiktlik/, is quite normal and even large

compared to the ranges often reported for /u/-related protrusion in the literature (Engstrand, 1981; Lubker & Gay, 1982; Perkell, 1986). More importantly, if a behavior is consistent (whether perceptible or not) it is necessary to incorporate it into any theory of how utterances are translated from phonological representation into a motor plan.

[7]One of our reviewers suggested that the peak in /iC$_n$i/ words might represent a return to a neutral position for the lips, from a retracted position for the preceding /i/ vowel. Although difficult to substantiate, this is a very plausible explanation. Note, however, that for the purpose of the argument, the origin of the peak in /iC$_n$i/ words does not matter—if a peak, or its effects, are present in contrastive contexts, then that peak represents a target that must be accounted for in the translation from the phonology to a motor plan.

[8]More extensive data on protrusion for all three intervocalic consonants are reported in Boyce (1988, 1990).

[9]One reviewer suggested the possibility that underspecification of vowels for the feature [nasal] might manifest itself as a smooth trajectory between flanking oral and nasal consonants with small perturbations due to intrinsic vowel height-velum height relationships. We would like to point out, however, that this hypothesis would not predict such similarity in movement traces as found, for example, in the early portions of /ə lasal/ and /ə lansal/.

[10]Keating also assigns narrow windows of velum height to consonants: one, in the low velum region, for nasal consonants and another, in the high velum region, for oral consonants. The problem with this proposal is that the relatively large size of the velum window for vowels (as compared to consonants) is derived from combining measures for vowels of different qualities. The question of whether a window for individual vowels is wider than that for individual consonants is an empirical one that has not yet been tested.

Haskins Laboratories Status Report on Speech Research
1990, SR-105/106, 153-172

Task Coordination in Human Prehension*

Patrick Haggard†

Movement patterns may be complex in the sense of being composed of separable component tasks. These components may be coordinated at some level by the voluntary motor system, in order to combine tasks into appropriate actions. This study describes the use of task interference methods and phase transition curves (PTCs) to quantify task interference in tasks that may have two components. Comparing the effects of task interference on the different components suggests how these may be coordinated during normal movements. These techniques can be applied to the coordination of hand transport and grasp aperture components in the reaching and grasping movements that people make in order to pick things up. Five subjects made cyclical movements which involved either composite reaching, or just the transport or grasp component in isolation, according to condition. The cyclical movements were "perturbed" by requiring a rapid transport or grasping response to an auditory signal by the contralateral hand. The pattern of phaseshifts, or changes in the timing of the cyclical task introduced by these "perturbations" was modelled using Phase Transition Curves, in order to assess the nature of the functional linkage between transport and aperture in normal prehensile movement. The results suggest a functional linkage between grasp aperture and hand transport in normal prehensile movement.

Functional Linkages at Task and Effector Level

There has been comparatively little interest, in the movement control literature, in complex movements which not only require coordination of multiple effectors, but also involve two or more component tasks. Yet many human actions seem to involve separable elements which are themselves tasks. Consider manual tracking, for example: skilled performance of this task requires both moving the eyes in pursuit of the target and also moving the manipulandum along the trajectory of the target. These operations use different effector systems, and are additionally distinct in that each is a task with its own immediate goals. However, there may also be some degree of "communication" or information-sharing between component tasks, in order to facilitate their fusion into a single action. This information-sharing, or functional linkage, is an important characteristic of coordinated movement.

In producing multi-task movements, the motor system may "parse" the overall goal into component tasks. In contrast with linguistic parsing (which decomposes a sentence into the smaller sections which together make up its meaning, and is hence purely analytic), motor task parsing is likely to be both analytic (complex movements must be decomposed into their components) and synthetic (the control of component tasks must be directed by the goals for the complex movement as a whole).

Methods for studying multi-task movements

A number of methods for identifying functional linkages of this kind have been proposed for multi-effector movements. These include finding invariant relationships between effectors' trajectories (Soechting & Lacquaniti, 1981), finding common temporal patterns between neural activations of

I am grateful to the Commonwealth Fund of New York for a Harkness Fellowship, which made this research possible, to Bruce Kay, Ignatius Mattingly, Elliot Saltzman, Michael Turvey, and the staff of Haskins Laboratories for advice and assistance, and to John Duncan, Ian Nimmo-Smith, Tim Shallice, Alan Wing and two anonymous reviewers for comments and discussion. The writing of this paper was supported by a Wellcome Trust Prize Studentship.

different effectors (Sears & Stagg, 1976), or finding stable movement patterns despite varying conditions of movement (Kelso, Saltzman, & Tuller, 1986; Kugler & Turvey, 1988).

This paper combines four distinct elements to create an analogous method for identifying functional linkages at the task level, as opposed to the effector level. First, a modified task interference or dual task paradigm is used to selectively disrupt the component tasks which contribute to a complex movement. Second, the primary task may be disrupted throughout its time course by applying a discrete, secondary task at various times. Third, the response of the primary task to interference is expressed as a function of the primary task phase using Phase Transition Curves (PTCs). Finally, the secondary task is considered to have a constant effect on the motor control system.

Task interference

Interference between two tasks has been held to reflect competition either for the capacity of a single, general-purpose central processing channel (Broadbent, 1958; 1982), or for specific cognitive resources or "modules" (Shallice, McLeond, & Lewis, 1985).

The logic of task interference studies requires extension to tackle the issue of multi-task movements. If a complex multi-task movement consists of two component tasks, then it should be possible to find secondary tasks which perturb each component. Where the component tasks are encapsulated, and do not communicate, it should be possible to disrupt each component task selectively without affecting the others. On the other hand, if component tasks share information, an interfering task aimed at disrupting one component task should also disrupt other components with which the first communicates. Thus, observing the behaviour of the system as a whole in response to selective perturbations can reveal the communication between component tasks. This method will be called Complex Task Interference (CTI).

When investigating the way in which the components of the primary task are combined in a normal complex movement, comparison of CTI and traditional, simple task interference (STI) results can be particularly valuable. Consider again the example of manual tracking. Suppose a secondary task, T, interferes with pursuit eye movements when these are performed in isolation (i.e., an STI condition), but does not interfere with moving the manipulandum through a trajectory. Now suppose T is performed during the course of a normal composite tracking movement (i.e., a CTI condition). Under these conditions, T may still interfere with eye movements, and may now also interfere with the kinematic trajectory of the hand, perhaps because this trajectory is planned on the basis of oculomotor information. Alternatively, if less intuitively, the interference between T and oculomotor pursuit observed in isolation may now diminish or disappear, perhaps because the simultaneous occurrence of the kinematic trajectory of the hand alters the way in which oculomotor pursuit is executed, leaving the trajectory of the hand unaffected. Thus, comparing the effects of various interfering tasks on two or more components of a complex task can be used to study normal coordination of those components.

The time course of the primary task

Previous dual task studies have generally used continuous tasks in both primary and secondary roles (see Welford, 1968, Table 4.2). But an alternative technique which measured the effect on a continuous primary task from a discrete secondary task would permit precise assessment of the control and processing required by the primary task throughout its time course. Interference from discrete secondary tasks has previously been used in this way in probe reaction time studies (Posner & Boies, 1971; Posner & Keele, 1968). Posner and Boies visually presented a warning signal, followed by a letter, and finally a second letter. In the basic experiment, both the warning interval and the inter-stimulus interval, which separated the appearance of the letters, were held constant. As the primary task, subjects were asked to judge whether the two letters were the same or different. The secondary task was a discrete response to a white noise probe, which could occur at one of several points in the sequence of primary task events. Posner and Boies found a systematic effect on the probe RT as a function of the point in the primary task sequence at which the probe occurred. Higher probe RTs were taken to reflect increased processing demands of the primary task, thus providing a history of the primary task's processing requirement.

Charting the time course of processing demands is a valuable adjunct to a task interference paradigm, since "time-sharing" may occur in many dual task situations. A continuous record of processing demands can indicate epochs where time-sharing is a more or less attractive strategy. In the case of movements, it may also be particularly valuable to compare the temporal pattern of processing with the kinematic pattern of the observed behaviour.

Phase transition curves (PTCs)

Continuous records of processing demands can also be achieved by considering the effects of a discrete secondary task on the primary task using Phase Transition Curves (PTCs).[1] PTCs describe the effects of a single perturbation on a stable rhythmic behaviour, using the observation that discrete perturbations can cause phaseshifts (temporal advances or delays) in a cyclic behaviour. PTCs represent the effects of the perturbation by an array of signed phaseshift values considered as a function of the phase in the cycle at which the perturbation occurs. The magnitude of the perturbing stimulus generally alters the pattern of phaseshifts, so many studies include stimulus magnitude as a parameter of the design. The effects on the cyclic behaviour as stimulus phase and stimulus magnitude are varied permit an assessment of the stability, and thus the control, of the behaviour.

The use of PTCs to analyse task interference involves modelling the system's response to perturbation in terms of its timing characteristics. While PTCs are a relatively new method in psychology, phaseshifts have often been noticed and discussed in the literature on human performance (e.g., the psychological refractory period; Kantowitz, 1974; Allport, Antonis, & Reynolds' "timing errors," 1972; Michon, 1966). Thus, the application of PTC methods seems promising.

To obtain the PTC, the phase of the cyclic behaviour at a given instant is defined as the time since a particular reference event occurred, expressed as the modulus of the mean period of the cyclic behaviour. Phase is normally measured from 0 to 1, with phases 0 and 1 being equivalent. Using the example of manual tracking again, one could ask subjects to track a cycling target, defining the point when the target passed a specified line on the screen travelling in a particular direction, say, as phase 0.

Perturbations are delivered at a variety of phases in the movement's cycle (see Figure 1).

The phase of the behaviour at the end of the perturbation is termed the "oldphase" (Kawato, 1981). The phaseshift is the temporal discrepancy between the phase of the observed post-perturbation waveform and the phase of a hypothetical post-perturbation waveform obtained by continuing the mean pre-perturbation waveform to infinity (the dashed line in Figure 1). Positive phaseshifts conventionally indicate a phase delay (the observed waveform lags behind the hypothetical waveform), and negative phaseshifts indicate a phase advance (the observed waveform precedes the hypothetical waveform). The phaseshift observed on the i^{th} cycle after perturbation is termed the i^{th} transient phaseshift.

The oldphase [ϕ] plus the i^{th} transient phaseshift [$\Delta\phi_i(\phi)$] equals the i^{th} transient newphase [$\phi'_i(\phi)$]:

$$\phi'_i(\phi) = \phi + \Delta\phi_i(\phi),$$

where (i=1...n).

If the effect of the perturbation is sampled at a suitable range of oldphases, the phase transition curve itself can be plotted in a plane whose axes are the oldphase and newphase, either by connecting all observed sets of coordinates of oldphase and i^{th} transient newphase, or by fitting a regression line to these coordinates (Figure 4 will offer an example of this). The resulting curve is called the ith transient PTC. PTCs differ from other performance measures in some important ways. First, each trial or observation contributes a single value for oldphase, and as many transient newphase values as may be measured. Thus, successive transient PTCs are not independent. Second, the representation of phases is "biperiodic": i.e., both abscissa and ordinate of the oldphase by newphase plane wrap around and repeat themselves continuously, since phase 0 and phase 1 are identical. Any datapoints that appear initially to be outliers may therefore be translated by adding or subtracting 1 to the oldphase or newphase coordinate, thus incorporating them into the rest of the distribution.

Some perturbations may cause a temporary phaseshift of the motion in comparison with its normal, unperturbed pattern. These responses to perturbation are captured by the transient PTCs. When the system has recovered from perturbation and returned to its original cycle period, the phase of the perturbed motion may also return to normal, indicating the system's complete recovery from perturbation, or the phase shift caused by the perturbation may persist indefinitely. The "steady state PTC" represents such long term effects of the perturbation on the behaviour. The experimental section of this paper, however, reports transient rather than steady state PTCs, since adjustments to perturbed movements are likely to have low latencies.

Once obtained, the PTC's shape and topology can elucidate the stability of the behaviour, and suggest the nature of its control. If the perturbation has no effect, the result is a phaseshift of zero: as the oldphase varies from 0 to

1, it is matched by the newphase to yield a PTC with gradient 1 and intercept 0. At the other extreme, if the perturbation is so disruptive that the cycle is stopped entirely, then the PTC will have a mean gradient of zero, and some newphase values will not be represented at all, since the cycle will begin again at a constant phase, returning to square one, as it were.

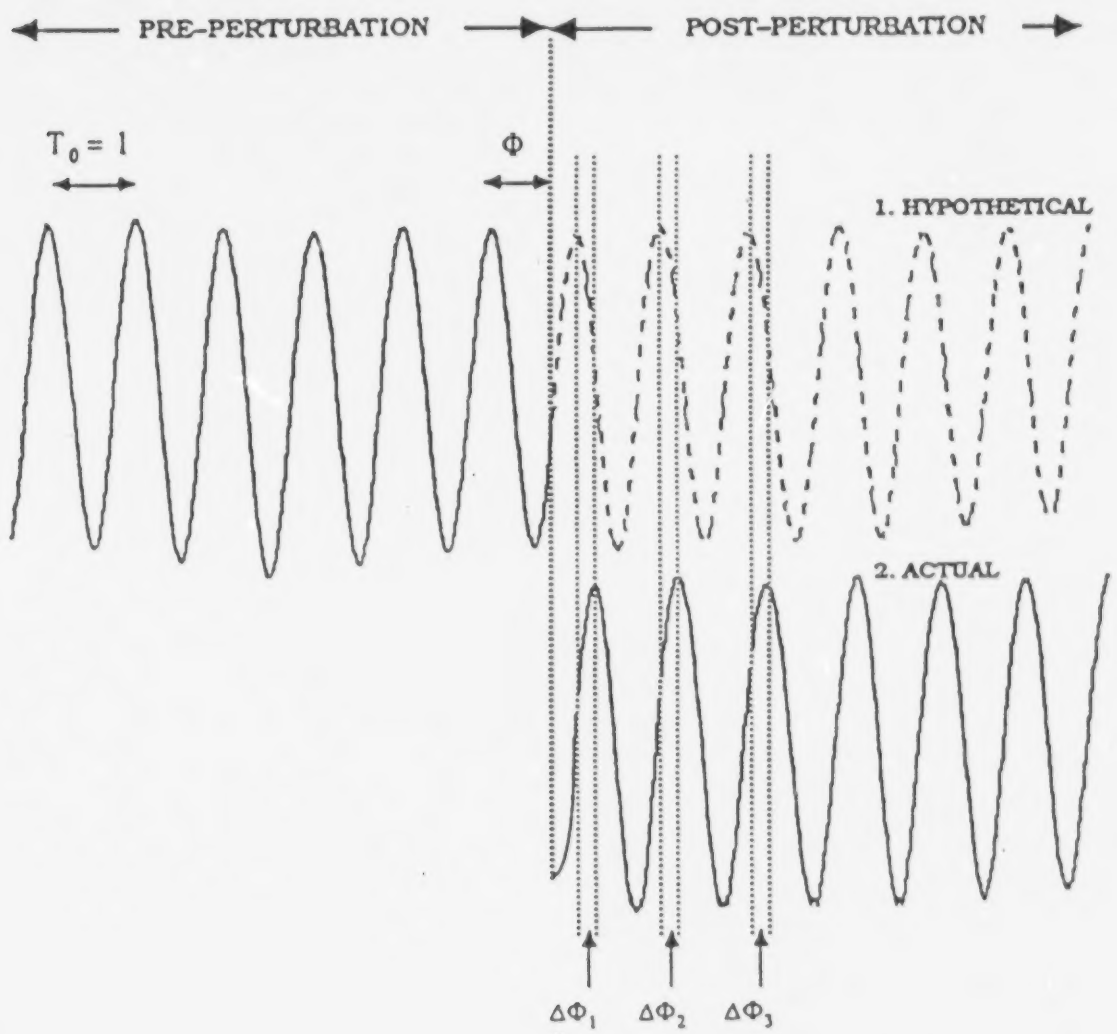


Figure 1. Definition of the Phase Transition Curve (PTC). First a reference event during the cycle is chosen, in this figure the peak of the waveform. All phase measurements are measured as moduli of the mean period of the pre-perturbation waveform: i.e., the mean pre-perturbation period, T_0, is assumed to be 1. The oldphase of the cycle is the phase of the cycle at which the perturbation ended. The first transient phaseshift is the temporal discrepancy between the actual post-perturbation waveform (solid line) and the pre-perturbation waveform hypothetically continued beyond the perturbation (dashed line) at the first reference event following the perturbation. As many transients as required may be measured. Larger (more disruptive) perturbations cause greater phaseshifts. The first transient newphase is the oldphase plus the first transient phaseshift. The first transient PTC is a fit to many datapoints representing the first transient newphase values plotted against the oldphase values. Perturbations may cause a simple linear delay or advance in the cyclic behaviour, or they may produce nonlinear phaseshifts whose magnitude and sign depend on the oldphase.

This may be described as a total "reset" of the behaviour. Other, intermediate degrees of resetting are possible, as when the perturbation causes the cycle to restart within a restricted range of newphase values which are nevertheless a systematic function of the oldphase, resulting in a PTC with a mean gradient of 0, but with locally nonzero gradients. Comparisons of actual PTCs with these two benchmark patterns can be used to quantify the effects of the perturbation delivered. In general, PTCs for small perturbations have mean gradients close to 1, and those for larger perturbations have mean gradients close to 0. All PTCs observed in biological systems have belonged to one of these two basic geometrical categories (Winfree, 1988).

PTCs thus provide a diachronic measure of stability, suitable for probing control throughout a continuous primary task. But there are also the following caveats: First, the less stable the studied rhythm is in the absence of perturbations, the less reliable PTC methods will be. Second, large numbers of observations are required to assess the effects of perturbations delivered throughout the cycle. Finally, the extensive development of the PTC technique by mathematical biologists has not yet produced a set of accepted statistical procedures either for fitting appropriate curves to sets of datapoints, or for testing hypotheses about the effects of perturbations using PTC results—this topic will be discussed later.

Applying PTCs to task interference

The early use of PTCs in investigating biological oscillators, and their present use in task interference differ importantly as regards the nature of the perturbing stimulus. The former tradition employs known, quantifiable stimuli, as when examining the effects of a flash of light on invertebrate circadian rhythms (Pittendrigh & Bruce, 1957), and precisely manipulates the stimulus strength to investigate what patterns of phase resetting of the underlying oscillator can be induced. In the case of task interference, however, the "strength" of the perturbatory task is not directly quantifiable, and is, in fact, the question being addressed by the investigation, rather than a planned parameter of the experimental design. Thus, task interference studies must infer the strength of the perturbation *post hoc* from the PTCs themselves, using a suitable statistical procedure.

One such procedure is as follows. First, a constrained linear fit of newphase values on oldphase values is performed. The slope of the regression line is constrained to be 1, and the intercept constrained to be 0. This regression line corresponds to the PTC that would be obtained with no perturbation. A second, less constrained fit is then performed to obtain a PTC which allows for any effect of the perturbation either by varying the slope or the intercept, or by adding varying nonlinear terms to the regression function, so that the newphase values may be more accurately predicted from the oldphase values.

An incremental F statistic (Kerlinger & Pedhazur, 1973) can then be used to test for a significant difference in the quality of the fits, adjusting for any differences in the number of terms used in the two regressions. A significant and positive incremental F means that the additional varying terms in the second regression have captured an effect of the perturbation on the timing of the behaviour. Significant and negative incremental Fs are also possible, since any additional terms may not improve the fit, while still using up degrees of freedom in the numerator. In this latter case, comparison of the two regressions would suggest that the perturbation had no effect. The same conclusion follows when the incremental F statistic is not significant. Use of the incremental F can be extended to compare the quality of fit between any two stages in the regression procedure: adding extra terms which genuinely capture the effects of the perturbation will produce a significant incremental F. The choice of which and how many terms are added to the regression to attempt to capture the effects of the perturbation is clearly of critical importance, and will be discussed later.

One comparison between regression fits is of particular theoretical interest. Comparing a linear fit to a second fit having both linear and nonlinear terms can separate the effects of the perturbation into a linear component, which is independent of the oldphase at which the perturbation occurs, and a nonlinear component, which is phase-specific. A linear effect could result either from a transient speeding up or slowing down of a central timekeeper, or from a fixed interval during which the activity was entirely suspended (in the case of a phase delay), after which it resumed normally. This would constitute a simple "dead time", rather than any substantial change in the behaviour. The nonlinear effect, however, involves a perturbation actually altering the form of the cycle in the phase plane, and thus represents a genuine change in the dynamics.

More specifically, linear phaseshifts may be due to the constant interval required to switch some

general attentional capacity between the responses required for each task.[2] This response-switching could be a part of a time-sharing strategy for combining tasks in a system with limited informational capacity (Broadbent, 1958; 1982). Nonlinear, phase-specific phaseshifts, on the other hand, may be due to more complex interference between tasks, such as disruption or degradation of pre-established sets of commands. Thus, comparing the nonlinear fit with a linear fit can capture the more interesting and dynamic effects of task interference, as opposed to the simple time-sharing ones mentioned by Broadbent.

Application to Human Prehensile Movement

The above methods can be applied to substantive issues in motor coordination. As an illustration, consider the combination of hand transport and grasping components involved in everyday prehensile movements, such as reaching out to pick up a glass of water. Three specific issues concerning these movements will be addressed: the separability of reaching and grasping into the operation of two movement systems, assessing the relation between these two processes, and the applicability of PTCs to this relationship.

Reach and grasp movements require both that the hand be transported to an appropriate point in space for contacting the object, and also that the grasp's opening and closing be appropriate for the intrinsic properties, such as size and mass, of the object to be picked up. They are thus both multi-effector and multi-task movements.

Jeannerod (1981) has proposed that these two components are performed by independent subsystems: an egocentric reaching component controlling the transport of the hand, and an object-centred component controlling the configuration of the fingers for grasping. He suggests that the two components are independent processes which are controlled separately, and do not share any information, except for a loose temporal coupling (Paulignan, Mackenzie, Marteniuk, & Jeannerod, 1990). Two strands of research seem to support this view: first, integrating egocentric and object-centred representations is a complex computational problem (Marr, 1982); second, signals for the control of grasp and of forearm position may be carried in separate tracts in the primate central nervous system (Lawrence & Kuypers, 1968). On the other hand, an architecture of two encapsulated processes is computationally less efficient and less flexible than a model in which the reach and grasp are coordinated to some degree and share information (e.g., Wallace & Weeks, 1988; Wing, Turton, & Fraser, 1986).

Relations between hand transport and grasp aperture

Selective task interference methods are well suited to testing Jeannerod's hypothesis of two independent processes. If Jeannerod's "two-process" hypothesis is correct, it should be possible to devise perturbations which affect the reach, but not the grasp, and *vice versa*. Furthermore, the effect of a given perturbation on each movement system should be the same whether the other component is operating concurrently or not; and the effects of a perturbation on the combined reach and grasp movement should be predictable from the perturbation's effects on the hand transport and grasp aperture systems in isolation, since the processes are assumed to be independent. For example, if a particular perturbation affects grasping behaviour, the two-process view predicts that it should have the same effect on isolated grasping, and also on grasping which is part of a composite reach and grasp movement. These predictions can be directly tested by comparing the results from complex task interference and simple task interference conditions.

In short, studying the differential effects of perturbatory tasks on the components in isolation and in the composite case should be informative about how the reach and grasp components are coordinated, and how, if at all, they share information. Furthermore, since timing changes are both known to occur in dual task performance involving motor control, and to be important to coordination (Allport et al., 1972; von Holst, 1973), PTC methods seem applicable to the coordination of hand transport and grasp aperture.

METHODS

Subjects sat comfortably in a dentists' chair which allowed unconstrained movements of both arms. The three primary tasks investigated were all cyclic, and all involved the right arm and hand only. They were:

1. Repeatedly grasping a dowel 28 mm in diameter with the right hand, without moving the arm.
2. Repeatedly reaching out with the right hand in the horizontal plane over a distance of about 30 cm, without opening or closing the hand.
3. Repeatedly reaching and grasping the dowel with the right arm and hand, as in a normal prehensile movement.

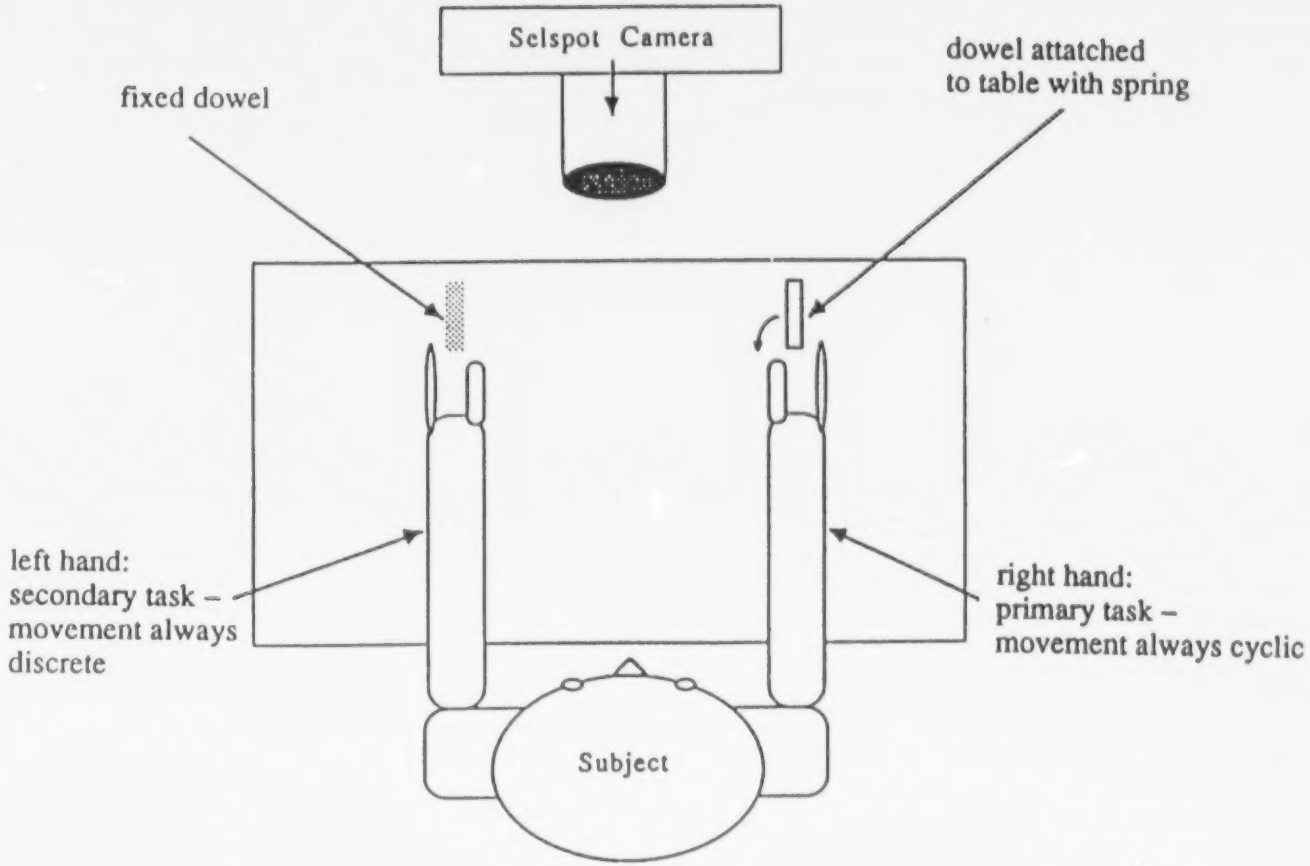


Figure 2. **Experimental apparatus.**

These primary tasks were chosen to involve, respectively, the grasping component alone, the hand transport or reaching component alone, and both hand transport and grasping together. The dowel used in the first and last tasks was located 30 cm in front of the start position, in line with the axis of the subject's right forearm. The dowel was fixed to the work surface with a spring, so it could be grasped, pulled and released, and would then spring back into the upright position. This arrangement ensured that the third condition, involving combined reaching and grasping movement, approximated normal prehension.

An auditory signal occurred at a random time during each trial. On hearing this signal, the subject performed one of two discrete secondary tasks with the left hand. One secondary task involved grasping a second dowel, again 28 mm. in diameter and positioned 30 cm in front of the subject in line with the left forearm. The other secondary task involved reaching out with the left hand in a horizontal plane in front of the body, over a distance of 30 cm. The two secondary tasks were chosen to be similar to the first two primary tasks hypothesized to involve the grasping component alone, and the hand transport component alone.

All combinations of the three primary tasks with the two secondary tasks were investigated, giving a total of six conditions. The data from three of these conditions are particularly relevant to the way in which grasping is coupled to hand transport in normal reaching. These three conditions, which are all reported here, involve the effect of a left hand grasp on right hand grasping, the effect of a left hand grasp on right hand reaching, and the effect of a left hand grasp on combined right hand reaching and grasping together. The three remaining conditions, involving the effect of left hand reach on each of the right hand primary tasks, do not directly address the relation of hand aperture to hand transport, and so are not reported here. An earlier pilot study also included a vocal response as a control secondary task; a condition analogous to Yamanishi, Kawato, and Suzuki's (1979) task B. However, since this condition, like that of Yamanishi et al. (1979) produced no phaseshifts at all, and made the experimental sessions unacceptably long, it was omitted from the final experiment.

Each condition was planned to include 48 trials, divided into four blocks of 12 trials each. All the trials in a particular block involved the same combination of a particular primary task with a particular secondary task. Subjects were instructed before each block which combination of primary and secondary tasks would be used in that block. One condition for one subject contained only 36 trials, due to error. Furthermore, a very small number of trials which seemed unusable (e.g., because the subject's movement accidentally took the hand outside the workspace) were repeated, so that each block contained at least twelve acceptable trials. The order of the blocks was randomized, with each subject receiving a different random sequence. Each trial lasted approximately 12 s, and was followed by an interval of a few seconds, during which the subject could stop the repetitive movement and relax, if she or he so wished.

Subjects were instructed to perform the primary task at a comfortable speed which allowed them to make the cyclical movement as regular as possible. They were further instructed to perform the secondary task as rapidly as possible on hearing the auditory signal, while attempting to keep the rhythm of the primary task regular. For the conditions involving grasping, subjects were instructed to grip the dowel between the pads of the thumb and the index finger. Practice was limited to two or three trials per subject, to check that both hands could make the full range of movements involved without being constrained by the apparatus. Since subjects were able to perform the cyclic primary tasks at an appropriate rhythm and to grasp the dowel almost immediately, further practice was judged unnecessary.

The movement of the thumb along the reach axis was monitored using a modified Selspot optical tracking system (accuracy 0.5 mm at 53 cm camera to subject distance: Edwards, 1985) positioned above the work surface, and parallel to it. The Selspot camera recorded in two dimensions the positions of Infra Red Emitting Diodes (IREDs) mounted on the distal interphalangeal joints of the right thumb and index finger. Since the thumb showed almost no side-to-side movement during reaching, and since its movement was parallel to the Selspot's y axis, movement in the y dimension of the IRED mounted on the thumb was taken as the transport component of the movement. The distance between the IREDs on the thumb and index finger in the orthogonal (x) dimension was taken as the grasp aperture component of the movement.

Kinematic data obtained from the Selspot can only give an approximate indication of when the right hand grasps the dowel. Therefore, a surface electrode was attached to the dorsal surface of the

subject's right hand, and another to the dowel grasped by the right hand. The electrodes were isolated from the main supply for safety. Changes in resistance between these two electrodes gave a precise indication of the time at which the subjects' right hand contacted and released the dowel.

The data were recorded on FM tape, digitized at 200 Hz, and subsequently processed on a computer. There were five subjects. All were right-handed, and were aged between 20 and 30. None had any history of neurological disorders.

Applying and fitting PTCs

Yamanishi et al. (1979) describe a procedure for finding PTCs by least squares regression fits to sets of data in oldphase and newphase coordinates.

PTC datapoints are phases, so any fit to these points must repeat or 'wrap around' on both oldphase and newphase axes, to capture the equivalence of phases across successive cycles. This requirement was satisfied by using a least squares regression with both linear and sinusoidal, nonlinear components. Yamanishi et al. (1979) used a single linear component: a gradient, which was fixed at 0 or 1, since these are the gradients underlying all observed PTCs. Their choice of gradient was made by taking the value which gave the best fit, though in other circumstances the choice could be made on *a priori* grounds such as magnitude of perturbation delivered. The present study added a further linear component not used by Yamanishi et al. (1979). This was an intercept which could vary freely. Including a linear intercept in the fitting procedure facilitated distinguishing between fixed delays or advances, and phase-dependent phaseshifts, as explained above.

The sinusoidal terms are taken in harmonic pairs from the Fourier series:

$$B_k \sin(2\pi kx) + C_k \cos(2\pi kx) \qquad (k=1...n)$$

where B_k and C_k were coefficients which could vary freely. Whereas Yamanishi et al. were primarily interested in the gross distinction between Type-0 PTC and Type-1 PTCs, more subtle distinctions can be made between fitted PTCs of the same type. Thus, the incremental F statistic can be used as described above to compare the quality of fit using just linear terms with the quality of fit using linear terms plus some sinusoidal, nonlinear terms intended to capture the effect of the perturbation. This statistic will be called the "*nonlinear vs. linear incremental* F." In this study, the fitting procedure was restricted to use only the first and second pairs of harmonics as the nonlinear terms, since these seemed to capture the pattern of phaseshifts in the data without consuming too many degrees of freedom in the regression.

Thus, the equation of the nonlinear regression fitted was:

$$y = Mx + A + B_1\sin(2\pi x) + C_1\cos(2\pi x) + B_2\sin(4\pi x) + C_2\cos(4\pi x)$$

where y = newphase, x = oldphase, and $M = 1$ (since the "perturbations" delivered were not highly disruptive). The first, second and third transient PTCs were calculated for each subject in each condition using this equation. Transient PTCs after the third were not calculated, because the movement generally returned to stable oscillation within three cycles. The issue of which nonlinear terms should be used is considered again in the methodological discussion.

RESULTS

Kinematics of waveforms

Figure 3 shows typical time series waveforms for the transport and aperture components of a right hand combined reaching and grasping movement during one trial.

The upper waveform represents the movement of the thumb in the y dimension, and can be taken as the hand transport component of the behaviour. The valleys of the upper waveform represent the starting position, while the peaks represent the position of the hand while grasping the dowel. The amplitude of the upper waveform, given by the valley-to-peak displacement, averages about 30 cm. The lower waveform represents the distance between the finger and thumb IREDs in the x dimension, and can be taken as the grasping component of the behaviour. The valleys of the lower waveform represent a closed grasp, while the peaks represent the maximum aperture attained just before grasping the dowel. The valley-to-peak displacement averages about 8 cm. The "notch" or horizontal portion just after maximum aperture represents the time during which the subject grasped the dowel with the right hand, while pulling it slightly backwards. Contact between the right hand and the dowel on the first cycle is marked by the first vertical dashed line.

The regular, cyclic nature of both waveforms can be clearly seen. Further, the two waveforms appear to be synchronized throughout the trial. The secondary task following the auditory signal or "perturbation," marked on Figure 3 by a further vertical dashed line, was a discrete left hand

grasp. The secondary task appears to have little effect on the timing of key events in the grasp waveform such as grasp contact (cf. Figure 7). However, the kinematics of the grasp waveform for this trial seem to show a slight increase in maximum grasp aperture immediately after the perturbation. While kinematic variations do not come within the scope of the present analysis using PTCs, the possibility of kinematic effects deserves further study.

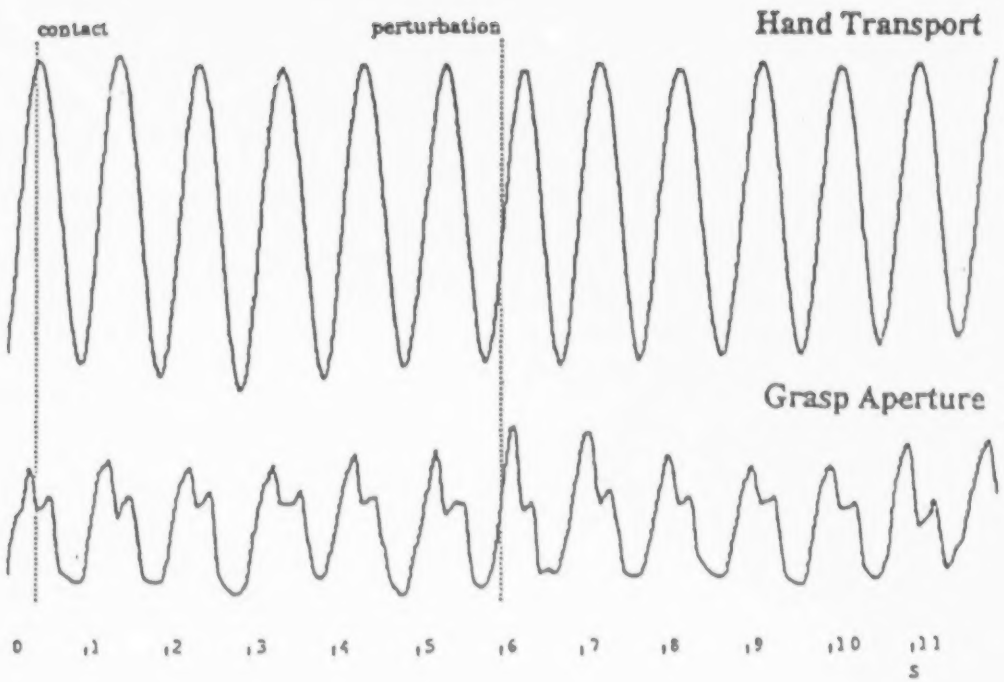


Figure 3. Typical time series waveforms for one trial in the right hand combined reaching and grasping condition. The upper trace shows the position in the y dimension of an Infra Red Emitting Diode (IRED) mounted on the distal joint of the right thumb. The lower trace shows the distance in the orthogonal, or *x*, dimension, between the IREDs mounted on the right thumb and the right finger. Note the clear periodic nature of both waveforms.

Phase Transition Curves

The results were analysed using PTCs, in order to assess the effects of the secondary tasks on the various primary tasks. Different events are used to define phase 0 for different experimental conditions. Where the primary task was repetitive right handed reaching, phase 0 represents the instant of the peak forward velocity of the right thumb as it approached the dowel. In these conditions, grasping the dowel generally occurs around phase 0.3, and the start of the forward transport of the hand towards the dowel generally occurs around phase 0.8. Where the primary task was right handed grasping alone, or combined reaching and grasping, phase 0 represents the instant of the subjects' hand contacting the dowel, as determined from the resistance between the electrodes on the hand and the dowel.

Figure 4 shows a typical set of datapoints, for subject 5 in the first condition: the effect of a left grasp perturbation on an isolated right handed grasp. The first, second and third transient newphase values are plotted against the oldphase values, with the first transient at the bottom of the graph, and the third at the top. Although the newphase is measured modulo 1, a constant of 1 has been added to all the second transient newphase values, and a constant of 2 to all the third transient newphase values, to separate each set of newphase values for display purposes. The solid curves, from bottom to top, are the first, second and third transient PTCs fitted to each set of datapoints using the equation given above. The dashed lines all have a gradient of 1 and, from bottom to top, intercepts of 0, 1, and 2. These are respectively the first, second and third transient PTCs which would have been obtained if the perturbatory task had had no effect on the timing of the primary task. The distance between the solid and the dashed line represents the overall effect of the perturbation, while the variations of curvature in the PTCs (solid lines) represent the nonlinear, phase-specific effect of the perturbation. More disruptive perturbations would tend to produce PTCs with a high mean displacement from the line of no perturbation, and with high curvature.

The data shown in Figure 4 are clearly better fitted by a PTC with an underlying gradient of 1 than by a PTC with a gradient of 0. On the other hand, the datapoints do deviate from a simple linear model with unit slope and zero intercept. This is particularly evident from the first transient of the data shown in Figure 4: note the concavity of the first transient PTC, and the displacement of the datapoints below the line of no perturbation around oldphase values 0.2 to 0.8. This contrasts with the much smaller displacements from the line of no perturbation for oldphase values from 0.8 to 0.2. If the requirement to make a left hand grasp occurs in the former section of the right hand's grasp cycle, the effect is a phase advance, whereas in the latter section of the cycle, the perturbatory task appears to have little or no effect. The effect of the discrete secondary task on the cyclic movement is thus clearly nonlinear, and the characteristics of the nonlinearities can be captured by the sinusoidal terms of the PTC.

The overall pattern of PTCs obtained is shown in Figures 5 to 7. The nonlinear vs. linear incremental Fs for the first transient PTCs shown in these figures are given in Table 1.

Taken together, the combination of simple and complex task interference results expressed in these figures suggests that the two-process theory of reaching and grasping must be incorrect.

Considering the first transient PTC only, the left handed grasp perturbation significantly disrupted right hand grasping in most subjects (see Figure 5).

In this figure, PTCs for subjects 1 to 5 are presented from left to right and from top to bottom. For each subject, the fitted PTCs and lines of no perturbation are arranged exactly as in Figure 4. Subject 5's data, already seen in Figure 4, are therefore shown again as the plot on the right in the bottom row. The individual datapoints, each representing one trial, are shown for the first transients, but have been omitted for the second and third transients, for purposes of clarity. When compared to a linear fit with gradient 1 and variable intercept, the nonlinear fits for the first transient PTC gave values for the incremental F statistic which were significant at the 5% level, indicating a significant effect of the perturbatory task, for each subject except subject 3. The nonlinear pattern is also clear from visual inspection: a perturbation occurring between oldphase values 0.2 and 0.8 caused substantial displacements from the line of no perturbation in all five subjects' data. Perturbations occurring between 0.8 and 0.2, by contrast, do not appear to produce large phaseshifts. For this condition, the "reference event" corresponding to phase 0 of the primary grasping task was the time of first contact between the subject's right hand and the dowel.

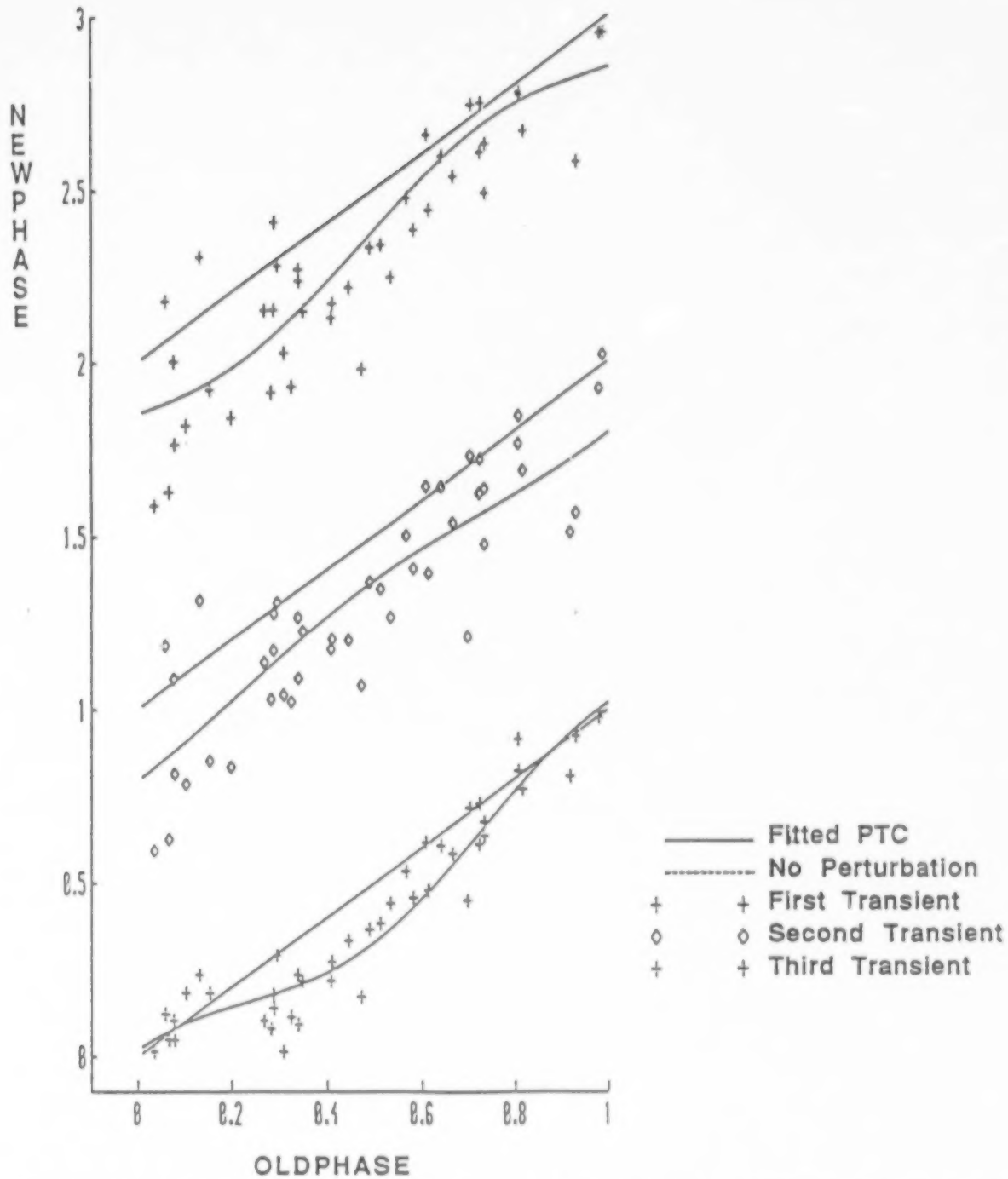


Figure 4. **A typical plot of, from bottom to top, oldphase against first, second and third transient newphases, together with the PTCs fitted to each set of datapoints (solid curves), lines of no perturbation (dashed lines), for subject 5 in the first condition, the effect of a left hand grasp perturbation on isolated right hand grasping. Although newphase values are measured modulo 1, a constant with value 1 has been added to all the second transient newphase values, and a constant with value 2 has been added to all the third transient newphase values for clarity of display. See text for further explanation.**

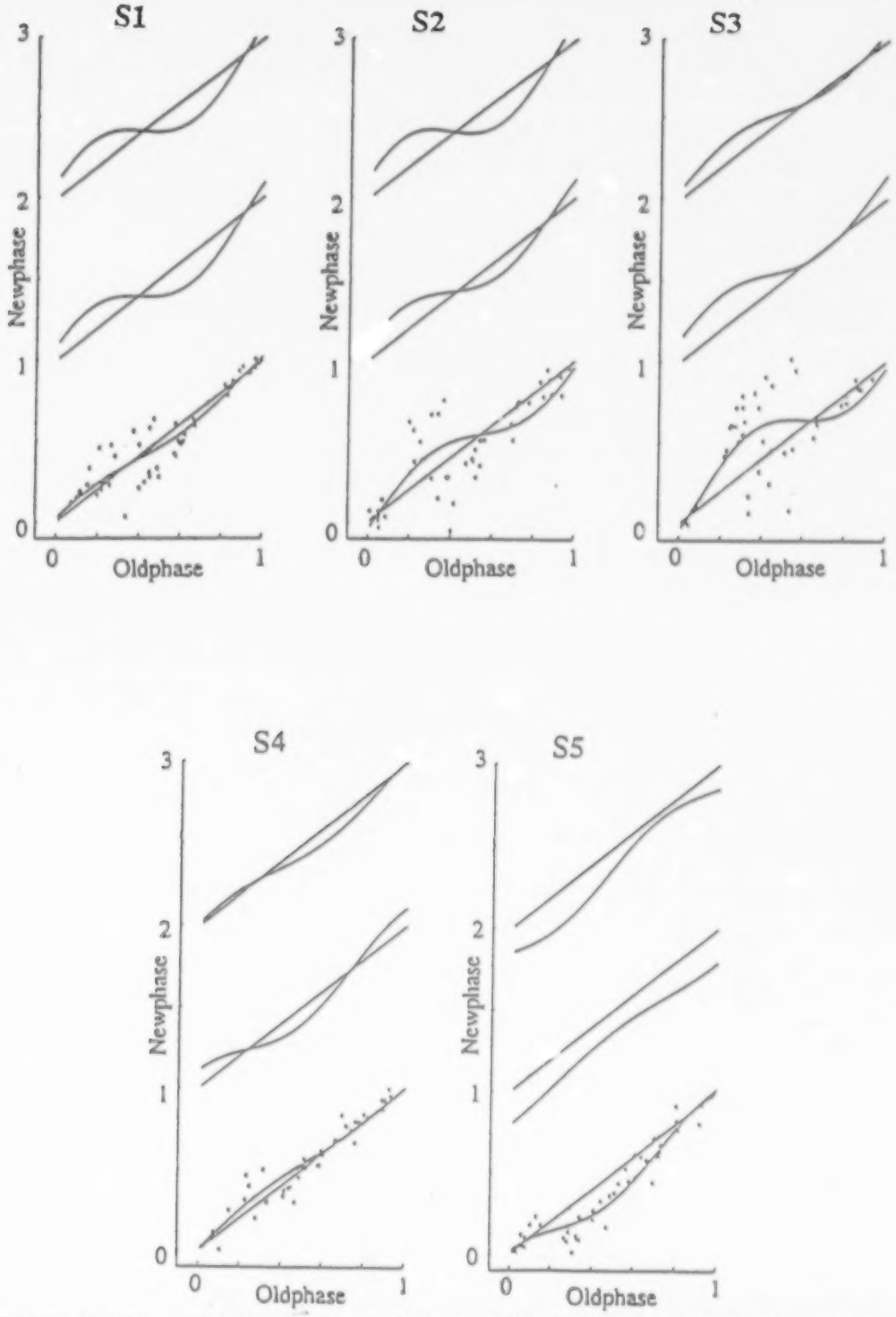


Figure 5. **PTCs for the effect of a left hand grasp perturbation on cyclic right hand grasping. The five graphs are, from left to right and top to bottom, for subjects 1 to 5. Each graph shows oldphase from 0 to 1 on the abscissa, and newphase on the ordinate. The three solid lines on each graph are, from bottom to top, the first, second and third transient PTCs. Thus, data from each trial contributes to each of the solid lines, since successive transient PTCs are not independent. The three dashed lines on each graph are lines with a gradient of 1 and an intercept of 0: these represent the PTC that would be obtained if the perturbation had no effect on the timing of the behaviour.**

Table 1. *Nonlinear vs. linear incremental Fs for first transient PTCs shown in Figures 5 - 7. See text for explanation.*

Corresponding Figure	Figure 5	Figure 6	Figure 7
Primary task (cyclic, right hand)	Grasp	Reach	Combined Reach and Grasp
Secondary task (discrete, left hand)	Grasp	Grasp	Grasp
Subject 1	Inc. $F(4,43) = 4.53$ $p < .001$	Inc. $F(4,55) = 1.37$ N.S.	Inc. $F(4,43) = 6.63$ $p < .001$
Subject 2	Inc. $F(4,42) = 7.48$ $p < .001$	Inc. $F(4,43) = 0.73$ N.S.	Inc. $F(4,31) = 2.03$ N.S.
Subject 3	Inc. $F(4,41) = 1.59$ N.S.	Inc. $F(4,[illegible]) = 1.83$ N.S.	Inc. $F(4,45) = 0.82$ N.S.
Subject 4	Inc. $F(4,38) = 2.83$ $p < .001$	Inc. $F(4,38) = 1.24$ N.S.	Inc. $F(4,38) = 1.84$ N.S.
Subject 5	Inc. $F(4,40) = 6.93$ $p < .001$	Inc. $F(4,43) = 0.85$ N.S.	Inc. $F(4,45) = 1.80$ N.S.

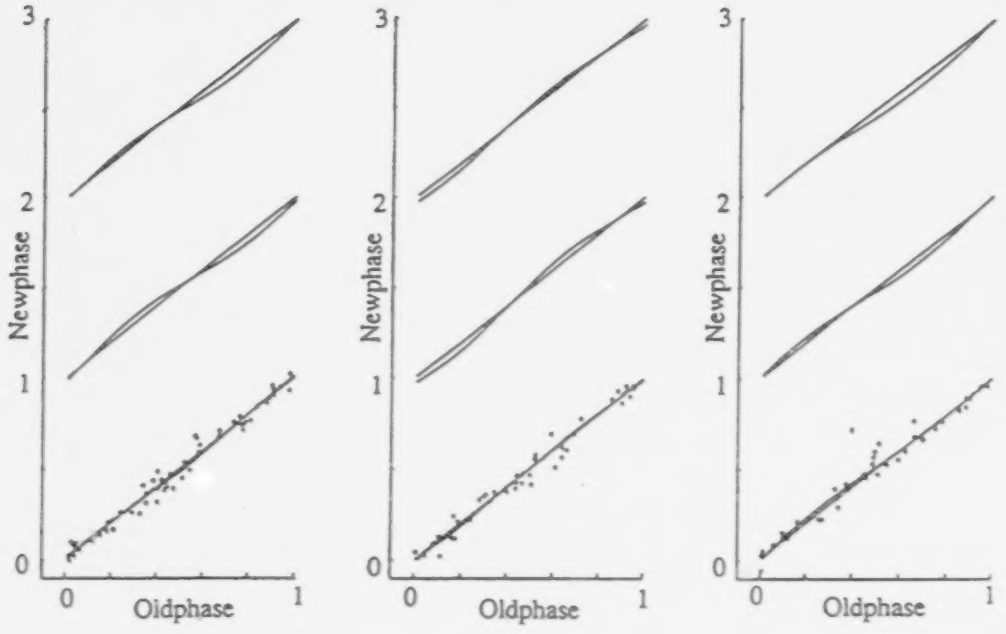
However, the same left hand grasp perturbation did not perturb right hand reaching as much (see Figure 6). None of the nonlinear vs. linear incremental Fs for the first transient PTCs in Figure 6 are significant, showing that the grasp perturbation did not disrupt the right hand's reach behaviour. The reference event corresponding to phase 0 in this condition was the peak velocity of the right thumb along the reach axis during the approach to the dowel.

These two results may be seen as an instance of like tasks interfering more than unlike tasks (cf. McLeod, 1977). However, such effects of task similarity are restricted to the case when both tasks are performed in isolation, and do not hold for complex reaching and grasping. In the case of complex movements, the two-process view predicts that the left grasp perturbation should perturb a right hand grasp even when the latter component is combined with a reach in a composite reach and grasp movement, since the effects of perturbations on the composite movement should derive straightforwardly from the effects on the component movements when performed in isolation. The PTCs of Figure 7 show that this is clearly not the case. As in figure 5, the reference event corresponding to phase 0 of the primary grasping task was the time of first contact between the subject's right hand and the dowel.

None of the nonlinear vs. linear incremental Fs for the first transient PTCs in Figure 7 are significant, except for subject 1. Thus, for most subjects, the left grasp perturbation did not severely disrupt the right hand's grasping when this was part of a composite reach and grasp movement. Note particularly that the first transient PTCs of Figure 7 exhibit much less curvature than those in Figure 5, and the datapoints are less scattered around the line of no perturbation, indicating that the effect of a left hand grasp perturbation on cyclical right hand grasping is less when the right hand's grasping movement is a component of a combined prehensile movement than when grasping is performed alone.

Discussion of Results

In the composite case, where the right-hand grasp is part of a normal reach-and-grasp behaviour, the grasp is no longer susceptible to perturbations that do have an influence when grasping is performed in isolation. The attempt at selective perturbation of the grasp aperture component alone therefore fails in the composite case. This suggests that the co-occurence of the right hand's reach causes a fundamental alteration in the way the right hand's grasping movement is controlled. Specifically, in the composite movement, the control of the grasp seems functionally dependent on the control of the reach. This result is inconsistent with any theory which proposes independent control of hand transport and grasp aperture.



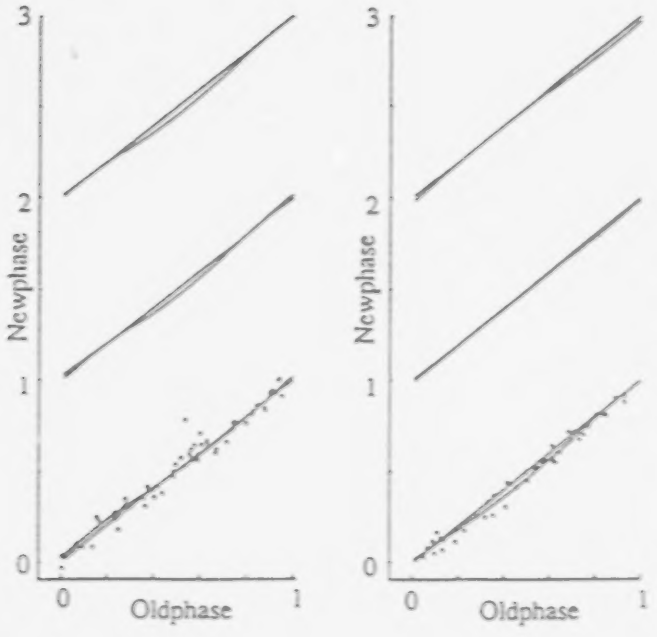


Figure 6. **PTCs for the effect of a left hand grasp perturbation on cyclic right hand reaching. The plot is arranged in the same way as Figure 5.**

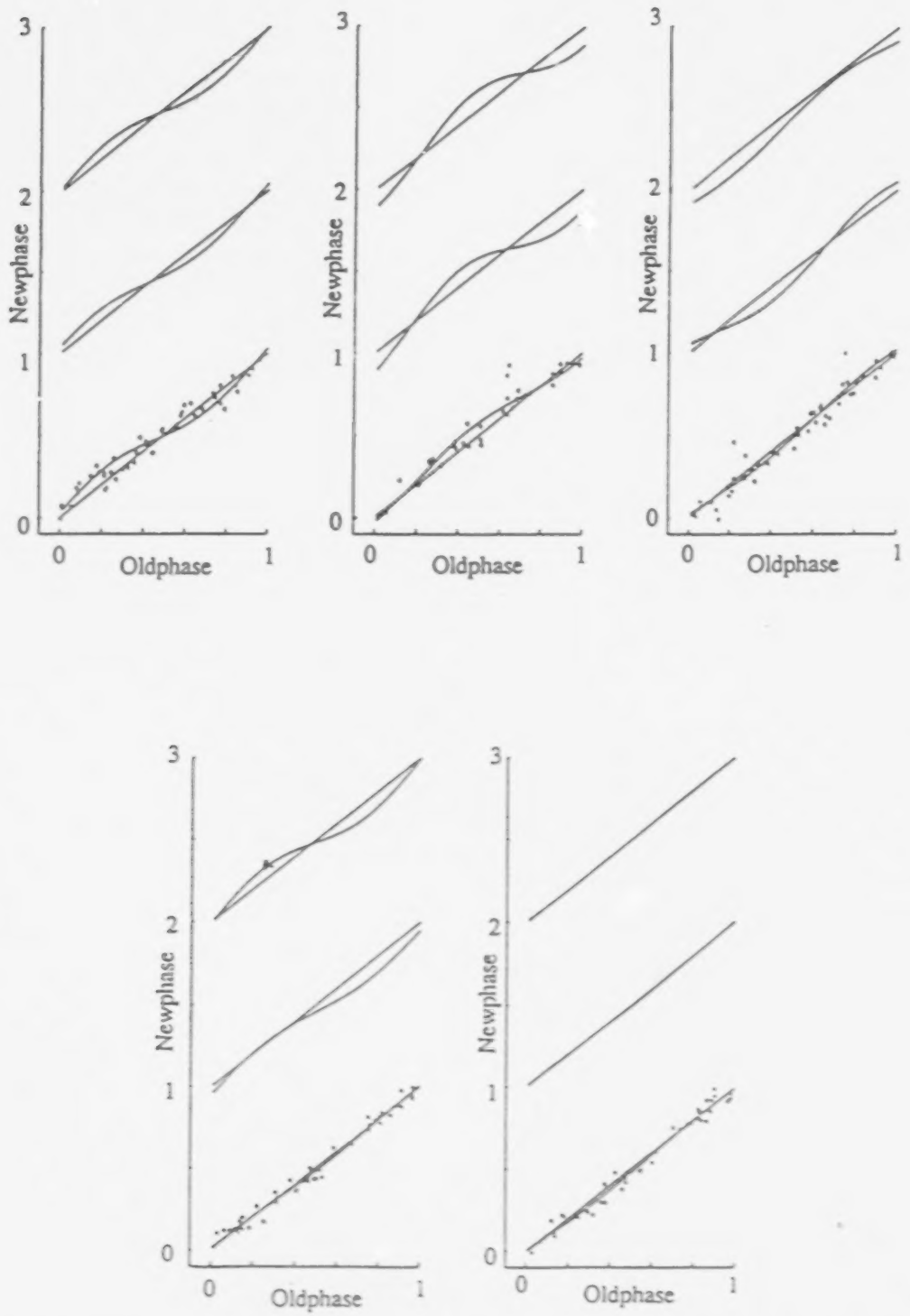


Figure 7. **PTCs for the effect of a left hand grasp perturbation on cyclic right hand grasping when the right hand grasping is part of a composite reach and grasp behaviour. The plot is arranged in the same way as Figure 5.**

An alternative interpretation of this result might attribute the lack of phaseshifts in Figure 7 to the fact that more time is available to organise the right hand's grasp when this is part of a composite reach and grasp movement, since in the complex, but not in the isolated condition, grasping takes only a proportion of the total time required for the movement. Thus, the grasp could be planned in advance, during this spare time, and the motor program simply run off at the appropriate instant. Adequate buffering of this kind might ensure that a discrete perturbation will not produce interference. Although this interpretation cannot be ruled out on the basis of the present data, it seems implausible for two reasons. First, pre-planning the grasp movement during any spare time before grasping actually occurs might produce an unacceptably large number of movements in which the grasp was not well adapted to the characteristics of the reach, and therefore of errors in grasping the dowel. Almost no such errors were observed. Second, in repetitive tasks such as those reported here, subjects would presumably tend to plan the grasp at approximately the same point in each composite movement. But, since perturbations were delivered randomly over all phases of the movement, a tendency for greater phaseshifts on trials where the perturbation occurred around this point would be expected. No such tendency is apparent in Figure 7.

The qualitative similarity of the first, second and third transient PTCs for each subject in each condition is also noteworthy. In particular, if a perturbatory task causes a phase-specific advance or delay (i.e., concavity or convexity of the PTC) for the first transient, the same pattern generally appears in the second and third transients, though perhaps to a slightly lesser extent. Thus, where a secondary task does disrupt the phase of the primary task, this disruption seems to be permanent: there is no systematic modulation of the timing of subsequent cycles of the primary task in order to return to the pre-perturbation phase. Rather, the perturbation permanently shifts the whole post-perturbation time series.

The differential penetrability of component tasks when interfering tasks are performed concurrently has been used to study the coordination of those components into a complex everyday movement. Since it is possible to reach without grasping and vice versa, the composite movement can presumably be "parsed" into a transport task and a grasp configuration task. However, these two components seem to be combined in a way that shares information, since the effect of a grasp perturbation on the composite movement cannot be predicted from the effects on the transport and aperture components independently. So a model which treats ordinary reaching and grasping as the addition of two entirely independent and encapsulated tasks is inadequate. Rather, the control of transport and aperture seems to be organised in a hierarchical, rather than a parallel fashion, since the co-occurrence of a reach with a grasp resulted in a pattern of perturbability which is similar to that for a reach, rather than some compromise between the patterns for hand transport and grasp.

Methodological Discussion

A combination of task interference and PTC methods has previously been used by Yamanishi et al. (1979). They obtained PTCs for the effects on cyclic finger tapping of three discrete, secondary tasks: a visual cognition task, a vocal reaction task, and a key pressing task performed with the contralateral hand. The key pressing task caused the greatest phaseshifts in the rhythm of finger tapping, and thus was judged most disruptive. Yamanishi et al.'s main interest was in the processes underlying timing control of the cyclic finger tapping movement of the dominant hand. By introducing a variety of concurrent tasks as "perturbations," they were able to introduce a resetting of the timekeeper or clock underlying finger tapping. This resetting was attributed to a change in processing, such as a delay in the neural mechanism responsible for timing control, caused by the perturbatory task. Whereas Yamanishi et al. investigated timing of a movement with only a single clear component, the present approach expands their methods to use PTCs and task interference to investigate coordination of multiple component tasks, such as the integration of hand transport and grasping in prehensile reaching movements, using selective perturbation techniques.

Possible effects on the Secondary Task

The effects of interference on the discrete, secondary task merit consideration. Yamanishi et al. showed that the reaction times in their secondary tasks were independent of the oldphase. They therefore argued that the phase-specific resetting of the cyclic tapping task could not be attributed to a simple pause required for the performance of the secondary task. The present study assumed, following Yamanishi et al., that reaction time for the discrete, secondary task

would constitute a phase-independent "dead time", producing a simple linear phase-resetting. Thus the nonlinear vs. linear incremental F statistic used automatically takes account of the reaction time component of the task interference by the discrete, secondary task, since the linear fit already includes an intercept term. In addition, the mean phase-resetting of the continuous, primary task introduced by the discrete, secondary tasks in this study was often very small (e.g., Figure 6), implying that the effects of interference on the timing of the discrete task may also be negligible.

The approach taken in this study, and in that of Yamanishi et al. contrasts with Posner and Boies' (1971) probe reaction time paradigm. Posner and Boies emphasised the performance of the discrete, secondary task, and ignored any disruption it might have caused to the continuous, primary task. Their justification is interesting in the current context: "In studying movements, it was found that the probe had little effect on the primary movement task, but that the probe RT reflected the central processing demands of the primary task in a very sensitive way." However, logical difficulties in using secondary task performance as a reliable and independent index of primary task control (Brown, 1968; Duncan, 1979) raise problems for probe reaction time studies. The alternative approach, as taken in this study, retains the concern with quantifying control throughout the movement, but emphasizes the effects of interference on the continuous, primary task, rather than on the discrete, secondary task. Thus, whereas Posner and Boies did not deliberately match primary and secondary tasks, the present technique specifically chooses discrete, secondary tasks with a view to their possible effects as selective perturbations of the primary movement.

Further issues regarding PTCs

A study by Kay (1986) used PTCs to assess the effects of a brief mechanical perturbation from a torque motor on human finger movements. Kay used the pattern of phaseshifts induced by the perturbation to investigate the dynamics of the oscillator presumed to underlie the movement. One problem in using mechanical perturbations may be the change in displacement of the effector caused by the torque motor, which can itself generate a reference event, thus artifactually producing an abnormally long or short "cycle," which is purely passive. Perhaps for this reason, Kay chose to use steady state PTCs only.

Both Yamanishi et al. and Kay presented their perturbations at phases of the movement (oldphases) which were determined by an on-line measuring device. Yamanishi et al. perturbed at oldphase values 0.0, 0.1... 0.9. Kay aimed to perturb at oldphase values 0.0, 0.25... 0.75, though he did subsequently calculate the precise oldphase from the movement waveform. This practice of perturbing only at or around particular oldphase values seems unfortunate in two respects. First, it makes perturbations statistically more predictable. Second, it has adverse consequences for curve-fitting. Clustering perturbations on specific oldphase values requires fitting a predominantly diagonal (type 1) or horizontal (type 0) PTC to a few vertical bars of datapoints. Not only is the detailed representation of phaseshifts in the regions between the bars entirely lost, but the results of fitting sinusoids to such "vertical" datasets can be deceptive. A statistically more acceptable approach is to deliver perturbations entirely randomly, without on-line monitoring and control of oldphase values. Where the number of trials is large, scatter of the data over all oldphase values is guaranteed. This latter approach is also technologically simpler.

The use by Kay and by Yamanishi et al. of steady state, rather than transient, PTCs raises two distinct problems. While steady state PTCs are better understood, their use implies discarding information about the behaviour's transient response to perturbation. For example, if the new-phase values for a particular set of oldphase values exhibit a pattern of oscillation between successive transients (such as a phase delay for the first transient, and a phase advance for the second), this would be equivalent to a negative autocorrelation at a lag of unity (Wing & Kristofferson, 1973). Since the first, second and third transient PTCs for the reaching and grasping data reported above are qualitatively similar, the processes responsible for the negative lag one autocorrelations found in other rhythmic movement time series, such as tapping, appear to be absent from the prehensile movements observed in this study. This difference merits further investigation.

Further, the precise time taken for return to the steady state could vary substantially from one trial to the next, and could also vary as some interesting function of the magnitude of the perturbing stimulus, or of the oldphase. This information, which may be relevant to the system's dynamics under perturbation (Scholz & Kelso, 1989), is similarly discarded when using steady state PTCs.

Finally, the problem of fitting PTCs deserves a few further comments. The regression procedure proposed by Yamanishi et al., which involve a linear plus sinusoidal least squares fit, clearly satisfy the requirement of a biperiodic regression. However, a number of difficulties remain. First, sinusoidal fits inevitably tend to obscure any interesting local discontinuities in PTCs. As such discontinuities have been found in some biological systems (though not yet in human movement), this may be an unfortunate omission. Second, each additional pair of regression terms taken from the Fourier expansion transfers two degrees of freedom from the denominator to the numerator. Thus, greater statistical significance for a fit to a set of datapoints may sometimes be achieved by reducing the number of sinusoidal terms. This is particularly the case where the number of datapoints is low. Fitting too few sinusoidal terms may yield significance at the price of failing to capture the actual pattern of phaseshifts in the data. On the other hand, using a larger number of sinusoidal parameters, thus including some of higher frequency, will give a more detailed fit to any confined regions on the abscissa where there are phaseshifts which are highly phase specific, in the sense that the response of the system to perturbation may be very highly dependent on the oldphase at which the perturbation is delivered. The interesting possibility of identifying such behaviour in human movement control encourages the use of a fair number of sinusoidal terms.

CONCLUSIONS

The methodological section of this paper has argued for the value of both phase transition curves and task interference techniques in studies of coordinated movement, and has discussed some modifications and extensions to these methods with a view to increasing their applicability. The experimental section has used these methods to suggest a functional linkage between grasp aperture and hand transport in human prehension.

REFERENCES

Allport, D.A., Antonis, B., & Reynolds, P. (1972). On the division of attention: A disproof of the single channel hypothesis. *Quarterly Journal of Experimental Psychology, 24*, 225-235.

Broadbent, D.E. (1958). *Perception and communication*. London: Pergamon.

Broadbent, D. E. (1982). Task combination and selective intake of information. *Acta Psychologica, 50*, 253-290.

Brown, I. D. (1968). Criticisms of time-sharing techniques for the measurements of perceptual-motor difficulty. In *Proceedings of the XVIth International Congress of Applied Psychology*. Amsterdam: Swets & Zeitlinger.

Duncan, J. (1979). Divided attention: The whole is more than the sum of its parts. *Journal of Experimental Psychology: Human Perception and Performance, 5*, 216-228.

Edwards, J. (1985). *Mandibular rotation and translation during speech*. Unpublished doctoral dissertation, City University of New York.

Holst, E. von (1973). *The behavioural physiology of animals and man: The collected papers of Erich von Holst, Vol. 1* (Tr. Robert Martin). London: Methuen.

Jeannerod, M. (1981). Intersegmental coordination during reaching at natural visual objects. In J. Long & A. Baddeley (Eds.), *Attention and performance IX*. Hillsdale, NJ: Lawrence Earlbaum Associates.

Kantowitz, B. H. (1974). Double stimulation. In B. H. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition*. Hillsdale, NJ: Lawrence Earlbaum Associates.

Kawato, M. (1981). Transient and steady state phase response curves of limit cycle oscillators. *Journal of Mathematical Biology, 12*, 13-30.

Kay, B. A. (1986). *Dynamic Modelling of rhythmic limb movements*. Unpublished doctoral dissertation, University of Connecticut.

Kelso, J. A. S. (1984). Phase transitions and critical behavior in human bimanual coordination. *American Journal of Physiology: Regulatory, Integrative and Comparative Physiology, 15*, R1000-1004.

Kelso, J. A. S., Saltzman, E., & Tuller, B. (1986). A dynamical perspective on speech production. *Journal of Phonetics, 14*, 29-61.

Kerlinger, F. N., & Pedhazur, E. J. (1973). *Multiple regression in behavioral research: Explanations and predictions*. New York: Holt, Rinehart & Winston

Kugler, P. N., & Turvey, M. T. (1987). *Information, natural law, and the self-assembly of rhythmic movement*. Hillsdale, NJ: Lawrence Earlbaum Associates.

Lawrence, D. G., & Kuypers, H. G. J. M. (1968). The functional organization of the motor system in the monkey. II. The effects of lesions of the descending brain-stem pathways. *Brain, 91*, 15-36.

Marr, D. (1982). *Vision*. San Francisco: Freeman.

Miall, R. C., Weir, D. J., & Stein, J. F. (1985). Visuomotor tracking with delayed visual feedback. *Neuroscience, 16*, 511-520.

McLeod, P. (1977). A dual task response modality effect: Support for multiprocessor models of attention. *Quarterly Journal of Experimental Psychology, 29*, 651-667.

Michon, J.A. (1966). Tapping regularity as a measure of perceptual motor load. *Ergonomics, 9*, 401-412.

Paulignan, Y., Mackenzie, C., Marteniuk, R. & Jeannerod, M., (1990). The coupling of arm and finger movements during prehension. *Experimental Brain Research, 79*, 431-435.

Pittendrigh, C. S., & Bruce, V. G. (1957). An oscillator model for biological clocks. In D. Rudnick (Ed.), *Rhythmic and synthetic processes in growth*. Princeton NJ: Princeton.

Posner, M. I., & Boies, S. J. (1971). Components of attention. *Psychological Review, 78*, 391-408.

Posner, M.I., & Keele, S. W. (1968). Attention demands of movement. In *Proceedings of the XVIth International Congressof Applied Psychology*. Amsterdam: Swets and Zeitlinger.

Scholz, J. P., & Kelso, J. A. S. (1989). A quantitative approach to understanding the formation and change of coordinated movement patterns. *Journal of Motor Behavior, 21*, 122-144.

Sears, T. A., & Stagg, D. (1976). Short-term synchronisation of intercostal motoneurone activity. *Journal of Physiology, 263*, 357-381.

Shallice, T., McLeod, P., & Lewis, K. (1985). Isolating cognitive modules with the dual-task paradigm: Are speech perception

and production separate processes. *Quarterly Journal of Experimental Psychology, 37A*, 507-532.

Soechting, J. F. & Lacquaniti, F. (1981). Invariant characteristics of a pointing movement in man. *Journal of Neuroscience, 1*, 710-720.

Wallace, S. A., & Weeks, D. L. (1988). Temporal constraints in the control of prehensile movements. *Journal of Motor Behaviour, 20*, 81-105.

Welford, A. T. (1968). *The fundamentals of skill*. London: Methuen.

Winfree, A. T. (1988). *Biological clocks*. Scientific American Monograph Series.

Wing, A. M., & Kristofferson, A. B. (1973). Response delays and the timing of discrete motor responses. *Perception and Psychophysics, 14*, 5-12.

Wing, A.M., Turton, A., & Fraser, C. (1986). Grasp size and accuracy of approach in reaching. *Journal of Motor Behaviour, 18*, 245-260.

Yamanishi, J., Kawato, M., & Suzuki, R. (1979). Studies on human finger tapping neural networks by phase transition curves. *Biological Cybernetics, 33*, 199-208.

Yoshizawa, M., & Takeda, H. (1988). An output regulation model of human input adaptability in the manual control system. *IEEE Transactions on Systems, Man and Cybernetics, 18*, 193-203.

FOOTNOTES

*A slightly different version of this paper appears in *Journal of Motor Behavior, 23(1)*, 25-37 (1991).

†Now at the Medical Research Council Applied Psychology Unit, 15 Chaucer Road, Cambridge, CB2-2EF, England.

[1]Phase transition curves are an analytical technique, rather than a behavioural observation, and bear no relation to the "phase transition" phenomenon observed by Kelso (1984) in human cyclical movement.

[2]Evidence from other fields confirms the linear nature of time-sharing and switching effects. Yoshizawa and Takeda (1988) accounted for performance on a continuous pursuit tracking task using "an inherent dead time included in the operator" Miall, Weir, and Stein (1985) also posited feedback delays to explain intermittencies in monkeys' visuomotor tracking.

Haskins Laboratories Status Report on Speech Research
1991, SR-105/106, 173-186

Masking and Stimulus Intensity Effects on Duplex Perception: A Confirmation of the Dissociation Between Speech and Nonspeech Modes*

Shlomo Bentin† and Virginia Mann††

Using the phenomenon of duplex perception, previous researchers have shown that certain manipulations affect the perception of formant transitions as speech but not their perception as non-speech 'chirps,' a dissociation which is consistent with the hypothesized distinction between speech and nonspeech modes of perception (Liberman, Isenberg, & Rakerd, 1981; Mann & Liberman, 1983). The present study supports this interpretation of duplex perception by showing the existence of a 'double dissociation' between the speech and 'chirp' percepts. In it, five experiments compared the effects of stimulus onset asynchrony, backward masking and transition intensity on the two sides of duplex percepts. It was found that certain manipulations penalize the chirp side but not the speech side whereas other manipulations had the opposite effect of penalizing the speech side but not the chirp side. In addition, although effects on the speech side of duplex percepts have appeared to be much the same as in the case of normal (electronically fused) speech stimuli, the present study discovered that manipulations which impaired the chirp side of duplex percepts had considerably less effect on the perception of isolated chirps. Thus it would seem that duplex perception makes chirp perception more vulnerable to the effects of stimulus degradation. Several explanations of the data are discussed, among them, the view that speech perception may take precedence over other forms of auditory perception (Mattingly & Liberman, in press; Whalen & Liberman, 1987).

When we listen to speech, we tend to be unaware of the auditory signal qualities that give rise to our linguistic percepts. Careful "analytic" listening can reveal many such qualities—the hiss of fricative noises, the pops and clicks of stop consonant release bursts, etc. (see Pilch, 1979; Repp, 1981). Yet there exist certain auditory qualities in speech that remain inaccessible to even the most analytic listener. These qualities reflect energy that resides in particular regions of the spectrum such as the frequencies of individual formants and their changes over time; they can be made audible only in certain experimental situations. For example, the 'chirpy' auditory quality of single formant transitions can be made audible when the transition is extracted from an utterance and presented in isolation. In that case the formant transition is heard as a "chirp," and discrimination between various pairs of transitions is a nearly continuous function of their difference in frequency modulation. Yet when the same transitions are integrated into an acoustic speech pattern, where they cue the distinction between certain stop consonants, their 'chirpy' quality is lost and discrimination is categorical (Mattingly, Liberman, Syrdal, & Halwes, 1971; Mann & Liberman, 1983).

A particularly appropriate method of comparing these two ways of perceiving formant transitions was devised by Rand (1974); its consequences

The order of authorship is alphabetic, to reflect the joint contribution of the authors. The study was conducted while SB was a visiting investigator and VM was a research associate at Haskins Laboratories; it was supported by NICHD Grant HD-01994 and BRS grant RR-05596 to Haskins. We thank Alvin Liberman for the valuable advice and continuous encouragement that he gave to this project, Bruno Repp for his help, and Hwei-Bing Lin for her assistance in conducting the last experiment.

have been dubbed 'duplex perception' (Liberman, 1979). To induce duplex perception, one starts with a minimal pair of acoustic speech stimuli which differ only in a single cue, /da/ and /ga/, for example. Each stimulus is then divided into two parts. One part is a critical cue for the distinction between the two syllables, e.g., the second or third formant transition. The other is the remainder of the stimulus, that portion which is the same for /da/ and /ga/ and which is referred to as the 'base.' By changing the harmonic structure of the transition (see Whalen & Liberman, 1987) or by presenting the base and transition to separate ears (see Liberman et al., 1981; Mann & Liberman, 1983, Nygaard & Eimas, in press) one then introduces a discrepancy that results in a new stimulus configuration in which listeners can infer that the two parts of the stimuli are produced by separate sources. When listeners hear stimuli in this new 'discrepant' configuration, two percepts arise: a speech sound and a chirp which seems to 'float' away from the speech. What is important about these two 'sides' of the duplex perception is that each involves perception of the formant transition. Listeners report /da/ or /ga/ according to the nature of the transition, indicating that they have integrated the transition and the base rather than having processed the base alone. They simultaneously hear a 'chirp' sound which arises from a separate spatial location than the speech sound and has the auditory quality of the transition presented in isolation—a quality which is not heard in the original /da/ and /ga/ stimuli.

Duplex perception offers a controlled way of comparing perception of formant transitions as part of speech and as non-speech, because the two percepts arise from the same physical stimulus configuration. The experimenter can hold all variables constant and selectively direct a listener's attention to either of two readily available and differently localized percepts. If the two percepts arise from the operation of different 'modes' of perception, as has been suggested by several authors (i.e., Liberman et al., 1981; Mann & Liberman, 1983; Repp, Milburn, & Ashkenas, 1983; Repp & Bentin, 1984), then it should be possible to separately and selectively alter each percept. If two separate modes exist, certain additions to the stimulus or changes in its structure might alter one but not the other.

Two previous studies of the duplex perception have confirmed one aspect of this prediction by showing that the chirp side of duplex percepts remains unaltered by manipulations that significantly alter the speech side. Liberman et al. (1981) found that prefixing the base with a noise appropriate to /s/ has no effect on listeners' ability to perceive /p/ and /t/ transitions in the other ear as rising vs. falling 'chirps.' However, the noise makes the transitions indistinguishable as cues for a speech percept, because it causes both /pa/ and /ta/ to be heard as /sa/. The effect of the /s/ noise on the speech side of the duplex percepts is exactly as it would have been had the transition and base been electronically fused. It reflects the operation of a "specifically phonetic process" (Liberman et al., 1981, p. 142) in which the perception of /p/ vs. /t/ is prevented by the fact that the /s/-noise has replaced the closure silence that is an important cue to the perception of stop consonants. Another study by Mann and Liberman (1983) makes a similar point. In that study, it was shown that preceding the base by natural tokens of the syllables /al/ and /ar/ has no effect on listeners' ability to discriminate the chirp side of duplex /da/ and /ga/ stimuli which contained transitions that varied in onset frequency. Yet the phonetic discrimination of these same transitions as cues for /d/ vs. /g/ is systematically influenced by the preceding syllables because they induce a change in the location of the category boundary. Here, as well, the influences on the speech side of duplex percepts are the same as those which occur when the base and transition are electronically fused; they presumably reflect a specifically phonetic process in which listeners take account of the assimilating consequences of coarticulating /l/ or /r/ with a following /d/ or /g/ (see Mann, 1986, for discussion).

Thus it seems evident that, by manipulating certain aspects of the stimulus, one can alter the speech side of duplex percepts while leaving the chirp side unchanged. However, a stronger case for the dissociation between speech and nonspeech perception requires a "double dissociation." It is desirable to show not only that the speech side of duplex percepts can be altered while leaving the chirp side unchanged, but also that the chirp side can be altered while leaving the speech side unchanged. To this end, we have examined the effects of various acoustic manipulations on each side of the duplex percept. Experiment I investigated the effect of a temporal separation between the base and the second formant transition. Experiments II through IV examined the effects of certain types of masking and Experiment V examined the effect of a decrease in transition amplitude. If any of these

manipulations has a greater effect on the chirp side than on the speech side, the double dissociation between two perceptual modes would be confirmed.

EXPERIMENT I

Experiment I examined the effect of a stimulus onset asynchrony (SOA) between the transition and the base in duplex stimuli, asking whether this manipulation will have selective effects on speech perception as compared to chirp perception. Given previous results (Cutting, 1976; Repp & Bentin, 1984) we had reason to think that speech perception would be penalized as separation of the base and transition approached 100 ms. However, previous research had not examined the effect of SOA on the chirp side of duplex percepts. It was important that we determine how each side is affected by SOA before we turned to our studies of the effect of backwards masking in Experiment II.

Method

Subjects

The subjects were 10 female undergraduates. Two additional subjects were excused after they failed to distinguish /ba/ and /ga/ in the duplex practice series that preceded the test series.

Materials

The duplex stimuli comprised a base and second- formant transitions that were adapted from two-formant synthetic approximations to the syllables /ba/ and /ga/, produced on the parallel resonance synthesizer at Haskins Laboratories. The base by itself sounded vaguely like "da"; it contained the first formant (F1) and the steady state of the second formant (F2). Its total duration was 300 ms, with a 25 ms amplitude ramp at onset, and a 100 ms amplitude ramp at offset. Its fundamental frequency decreased linearly from 114 to 79 Hz. During the first 50 ms, F1 rose from 100 to 765 Hz, at which point it became steady-state and was joined by a constant F2 at 1230 Hz. There was no energy in the F2 region during the first 50 ms. The F2 transitions were synthesized separately from the base, with pitch and amplitude contour identical to that of the first 50 ms of the base. The /ba/ transition started at 924 Hz and rose linearly to 1230 Hz; the /ga/ transition started at 2298 and fell linearly to 1230 Hz. The absolute amplitudes of the transitions were set at the values F2 would have had in intact syllables.

To ensure that subjects could adequately perceive the chirp and speech components of the duplex stimuli, three practice sequences were recorded. The first was designed to familiarize subjects with the /ba/ and /ga/ syllables. It presented the base electronically fused with each of the F2 transitions, five times each and then five times in alternation. The second practice series was designed to familiarize the subjects with the chirps; it presented the isolated F2 transitions five times each and then five times in alternation. The third practice series familiarized subjects with duplex percepts; it presented the base and F2 on separate channels in onset synchrony, with each transition heard five times separately and then five times in alternation. In the test series designed to assess the effects of SOA, the /ba/ and /ga/ transitions preceded the base eight times at each of eight different SOA's: 0, 20, 40, 60, 70, 80, 90 and 100 ms. This yielded a total of 128 stimuli that were recorded in random sequence with inter-trial intervals (ITI's) of 2.5 s and longer pauses between blocks of 16 stimuli.

Procedure

Each subject participated in two sessions, counterbalanced across subjects. Here, and in all the other experiments reported in this paper the base was always heard in the left ear, the second formant transitions were always heard in the right ear (in our informal experience, ear differences tend to be negligible). One session required labeling of the speech percepts as "ba" and "ga"; it was preceded by the first and third practice series (i.e., the fused syllables and the duplex stimuli). The other session required labeling of the chirps as "rising" or "falling"; it was preceded by the second and third practice series (i.e., the isolated F2 transitions and the duplex stimuli).

Results and Discussion

As SOA increased, the identification of the speech percepts as "ba" or "ga" became considerably less accurate. In contrast, identification of the nonspeech percepts as "rising" or "falling" chirps improved slightly. These results appear in Figure 1, where it may be seen that speech identification declined with SOA, from a level of almost 95% correct to a level of 60% accuracy as SOA approached 100 ms. In contrast the accuracy of chirp identification increased from a level of 85% to a level of 95% accuracy for SOA's greater than 0 ms. A repeated-measure two-way ANOVA with the factors Task (speech or nonspeech identification) and SOA confirmed these observations. There was a significant effect of Task ($F(1,9)= 12.39$, $MSe= 780$, $p<.005$) and of SOA ($F(7,63)=3.39$, $MSe=65$, $p<.004$). More

importantly, there was a significant Task by SOA interaction ($F(7,63)=8.75$, $MSe=73$, $p<.0001$).

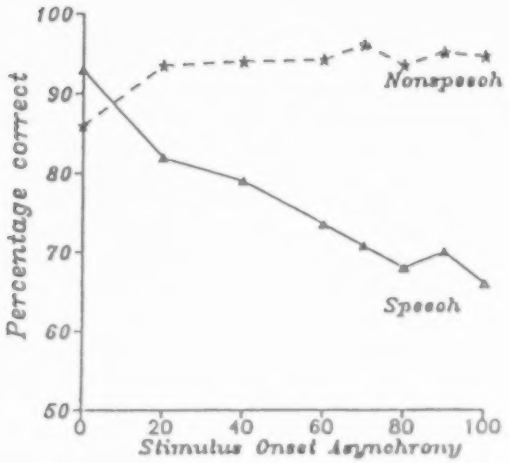


Figure 1. **Categorization of speech and nonspeech sides of duplex percepts at different SOA's (in ms) between the transition and the base.**

These results indicate that perception on the speech side of the duplex percepts is indeed disrupted by a temporal separation between base and transition. The chirp side of duplex percepts was not impaired by the temporal separation—it was slightly facilitated—and this is consistent with the claim that chirp and speech perception in the duplex phenomenon are mediated by two different modes of perception, a claim which will be more rigorously tested in Experiments II-V. The primary contribution of this experiment is its confirmation that the duplex speech percepts rest upon integration between the base and second formant transition. Although listeners have been reported to be capable of labeling isolated second formant transitions as speech (Nusbaum, Schwab, & Sawusch, 1983), it appears that integration of the transitions and base is an important component of the speech percepts in the duplex phenomenon that concerns us here.

EXPERIMENT II

In this experiment, we sought to further dissociate speech and chirp perception by finding a manipulation which alters chirp perception but leaves speech perception unaffected. We chose to examine the effects of a white noise presented immediately after the transition in the same ear. Our test involved a partial replication of Experiment I in which we examined the effect of increasing SOA, while placing white noise following the transition. The possibility that a backwards mask might have selective effects on the chirp side of duplex percepts was suggested by the results of Experiment I, where chirp perception improved slightly as presentation of the base was delayed in time. We were also prompted to consider the possibility that backward masking might have a greater effect on chirp perception than on speech perception by some of the data in Mann and Liberman (1983). Their first experiment on duplex perception revealed that subjects' ability to discriminate chirps in the duplex condition was inferior to their ability to discriminate chirps in isolation, and it was speculated that the decrease in chirp performance in the duplex condition was a consequence of the distracting circumstance of hearing two simultaneous percepts. However, we noted that any penalizing effects of the "distraction" were unique to chirp perception, hence selective masking seemed as reasonable an account as selective attention.

Method

Subjects

The same subjects that were tested in Experiment

Materials

The stimuli were identical to those used in Experiment I with the exception that a masking stimulus immediately followed each of the two transitions, presented to the same ear. The masking stimulus was a 15 ms burst of white noise with abrupt on- and off-sets; intensity was slightly above the peak amplitude of the isolated transitions. In the test series, the interval between transition and noise was fixed at 0 ms, and each transition preceded the base eight times at each of three different intervals: 0, 20, and 40 ms. This yielded a total of 48 stimuli that were recorded in random series with the same ITI's etc. as in Experiment I.

Procedure

The test series were presented immediately following Experiment I. Thus, before being given the present test, all subjects were exposed to the three practice and two test series employed in Experiment I. All subjects were tested twice: In one session they were instructed to direct attention to the speech side of the duplex and categorize the syllables as /ba/ or /ga/. In the other, they were asked to attend to the nonspeech side and to categorize the chirps as "rising" or "falling." The order of the tasks was counterbalanced across subjects.

Results and Discussion

Because the same subjects and stimuli had been employed in Experiments I and II, the effect of the white-noise backward mask on perception of speech and chirps at each SOA was assessed in comparison to the unmasked condition in Experiment I. These results are presented in Figure 2 where the solid lines represent data obtained in Experiment I ('unmasked' condition) and the dotted lines represent those obtained in Experiment II ('masked' condition). Speech perception is on the left, and chirp perception on the right.

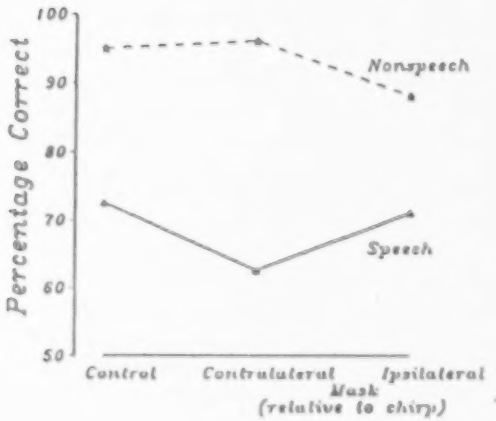


Figure 2. **Backward masking effects on speech and nonspeech sides of duplex percepts at different SOA's (in ms) between the transition and the base.**

The differential effect of the white noise mask on the perception of speech and of chirps is evident: The presence of the mask had a substantial interfering effect on chirp perception, but no effect on speech perception. A repeated-measures three-way ANOVA with the factors Task (speech or chirp), SOA, and Masking confirmed these observations. As in the previous experiment, the effects of task, of SOA, and the interaction between them were significant ($F(1,9)=12.39$, $MSe=780$, $p<.005$; $F(7,63)=3.39$, $MSe=65$, $p<.004$, and $F(7,63)=8.75$, $MSe=73$, $p<.004$ respectively). The effect of the noise mask was significant $F(1,9)=7.59$, $MSe=269$, $p<.02$, and the Task by Masking interaction was highly significant, $F(1,9)=53.47$, $MSe=72$, $p<.0001$.

The present results would seem to complement previous studies which showed selective effects on the speech percepts (Liberman et al., 1981; Mann & Liberman, 1983), offering a case of selective effect on chirp perception which supports the conclusion that the two sides of the duplex phenomenon represent the operation of two different modes of perception. But further research is required to discover how the effects of the mask is to be explained.

EXPERIMENT III

One explanation of the differential effect of the white-noise mask on chirp and speech perception is that the mask involved mechanisms that operates only within a particular auditory 'stream.' Drawing on the work of Bregman (1978; 1981) we might suppose that in the processing of auditory signals, 'scene analysis' incorporates the transition into two separate 'streams'. On the basis of spatial location (i.e., ear of origin) the transition and base would be assigned to separate 'streams' whereas on the basis of a common acoustic attribute—the F0 contour—the base and chirp would be assigned to one and the same stream. We might then postulate that, if masking is stream-specific, masking of the chirp side of duplex percepts should decrease when the white-noise mask is presented to the ear receiving the base (i.e., there should be less masking because the mask and chirp are now perceived in separate 'streams'). Another possibility is that speech perception is more tolerant than chirp perception of the signal degradation induced by a masking stimulus. If this is so, a contralateral mask should have the same effect as the ipsilateral mask employed in Experiment II, masking the chirp side of the duplex percepts more than the speech side. This would be consistent with Massaro's (1970) report that contralateral backwards masking interferes as much with pitch (i.e., chirp) perception as does binaural backward masking. As a test of these hypotheses, Experiment III studied the effects of a noise mask placed either contralateral or ipsilateral to the transition. A new set of stimuli modeled after those of Repp and Bentin (1984) was used so that we might establish the generality of Experiment II to stimuli which contained third formant transitions.

Method

Subjects

The subjects were twelve undergraduates (three male, nine female) who had participated in a pilot experiment which used the same duplex /da/- /ga/ stimuli. They had each demonstrated themselves capable of hearing the speech and chirp percepts in these stimuli.

Stimuli

The stimuli for this experiment were three-formant approximations of /da/ and /ga/, newly synthesized on the Haskins Laboratories equipment, and a 15 ms burst of white noise. In contrast to the earlier /ba/ - /ga/ stimuli, the critical distinction between the present syllables was carried by the F3 transition. The base, which sounded like "da" in isolation, was 250 ms in duration with a 50 ms amplitude ramp at onset and a constant fundamental frequency of 100 Hz for the first 100 ms followed by a linear decrease to 80 Hz at offset. F1 began at 279 Hz and increased linearly in frequency during the first 50 ms to a steady-state of 765 Hz. F2 began at 1650 Hz and decreased linearly during the first 50 ms to a steady-state of 1230 Hz. No steady-state F3 was included, as Repp and Bentin (1984) had shown that this was not critical and made the base sound more like "da." The /da/ and /ga/ F3 transitions were each 20 ms long, the short length being designed to enhance the effects of masking (the relative effectiveness of such short transitions had also been shown in Repp & Bentin, 1984). The /da/ transition began at 2800 Hz and decreased to 2745 Hz, the /ga/ transition began at 1800 and increased to 1945 Hz.

Three test sequences were prepared, one control (unmasked) and two masking conditions differing in the relative positions of the transition and the noise mask. In the ipsilateral masking condition, which replicated the masking condition of Experiment II, the 15 ms noise burst immediately followed the transition and was heard in the same ear. In the contralateral masking condition, the mask occurred in the same temporal position, but was heard in the same ear as the base. SOA between the transition and base was set at 35 ms to avoid actual overlap of the mask with either the base or the transition. The control condition presented the transition and the base without the mask, at 35 ms SOA. Each sequence comprised 36 /da/ and 36 /ga/ trials, randomized into three blocks of 24 trials each. ITI was 2.5 s and there were longer pauses between blocks of stimuli.

Procedure

Each subject participated in a speech and nonspeech session, with order counterbalanced across subjects. Speech percepts were labeled as "da" or "ga," and the labels 'high' and 'low' were used for the chirps, following Repp and Bentin (1984).

Results and Discussion

Consistent with the SOA effects observed in Experiment I, chirp identification was superior to speech identification in all three conditions (Figure 3). The more important result, however, was that the effect of the mask on perception of the transition depended on whether it was heard in the same ear as the transition, or the same ear as the base. When the mask was in the the same ear as the transition ("ipsilateral" condition) it reduced the accuracy of chirp identification, but had little effect on the accuracy of speech identification, replicating Experiment II and extending that result to third-formant transitions. In contrast, when the mask was in the same ear as the base ("contralateral" condition), it penalized speech perception but had no effect on chirp perception.

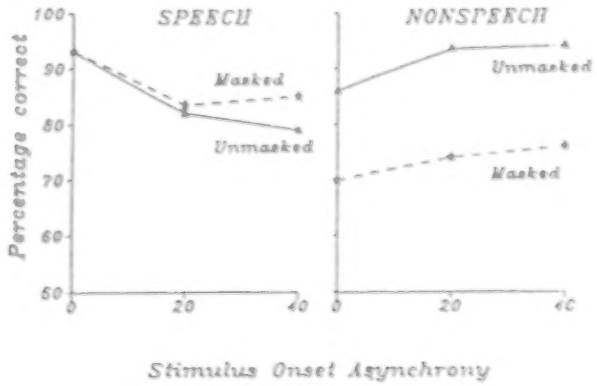


Figure 3. The effect of ipsilateral and contralateral backward masking on the speech and nonspeech sides of duplex percepts.

These observations were confirmed by a two-way repeated- measures ANOVA conducted with the factors Mode (speech, nonspeech) and Masking condition (control, ipsilateral, contralateral). The effect of Mode was significant, $F(1,11)=26.79$, $p<.0004$, and Mode and Condition interacted, $F(1,15)=5.49$, $p<.025$. Tukey HSD post-hoc comparisons revealed that the labeling of the speech percepts was significantly worse in the contralateral masking condition than in either of the other two conditions ($p<.01$). In contrast, labeling of the chirps was significantly worse in the ipsilateral masking condition relative to the other two conditions ($p<.01$).

It appears that a noise which occurs in the same ear as the base interferes with the speech side of duplex percepts but has no particular effect on chirp side, whereas a noise in the same ear as the transition has the opposite effect. These results are consistent with a account in which masking is 'stream'-specific and we shall defer further discussion of the role of scene analysis in duplex perception until the final discussion. The fact that the contralateral mask had a selective effect on speech perception does not support a view that speech perception is inherently more tolerant of degraded signals. The fact that the contralateral mask had less of an effect on chirp perception than the ipsilateral mask might further seem to be at odds with Massaro's (1970) report that contralateral backward masking of tonal targets by tonal masks interferes with pitch perception as much as does binaural backward masking, but it should be remembered that our stimuli are considerably different from his, as ours involved brief white-nose masks and duplex stimuli in which speech and non-speech percepts are simultaneously present.

EXPERIMENT IV

In Experiments II and III we had observed that ipsilateral backward masking had a greater effect on the chirp side of duplex percepts than on the speech side. We now turn to the question of whether this type of masking influences chirp perception in the case of isolated formant transitions. A control of this sort offers a test of a 'naive' auditory masking account in which masking should be more-or-less the same in duplex stimuli and in isolated transitions. Experiment IV addressed this possibility, using several mask intensities in order to maximize the possibility of finding masking in each condition.

Method

Subjects

The subjects were 12 Yale undergraduates who served as paid volunteers. They were screened from a larger pool of Yale undergraduates, using a selection criterion of > 80% correct identification of syllables in duplex presentation. The data of one subject were excluded from analysis because of very poor performance in identifying isolated transitions as chirps, which left 11 subjects.

Stimuli

The duplex stimuli were derived from the same synthetic /da/ and /ga/ syllables that had been used in Experiment III, except that there was no SOA between the base and the transitions. The masking noise employed in this experiment was excerpted from white noise digitized at 10 khz and low-pass filtered at 4.9 khz. It was 10 ms in duration and at its highest amplitude the RMS intensity was about 32 dB relative to that of the transitions. Four additional masks were created by digitally attenuating the noise in steps of 6 dB.

Procedure

The five mask intensity conditions and the "no mask" base-line condition were presented in a blocked design. Each block contained 48 stimuli, 24 /da/ and 24 /ga/ presented in random order. The blocks were presented in a fixed order for all subjects, the first block was the base-line condition, and the next five blocks presented masked stimuli with the relative intensity of the mask increased in equal steps of 6 dB from 8 dB (block 2) to 32 dB (block 6). Subjects were tested in two sessions, one for speech and one for nonspeech perception, and the order was counterbalanced across subjects. In the non-speech condition, subjects listened to the tape twice, once when the bases were presented to the ear opposite the transitions (i.e., the normal duplex listening condition) and once when the channel presenting the bases was disconnected such that only the isolated transitions could be heard. The order of the two non-speech conditions was also counterbalanced.

Results and Discussion

The results for the three different conditions are compared in Figure 4, where the mean accuracy of identification in each condition appears as a function of mask intensity. Identification of the speech and chirps in the baseline (no-mask) condition was uniformly high, but addition of the masking noise markedly reduced performance for chirps in the duplex condition.

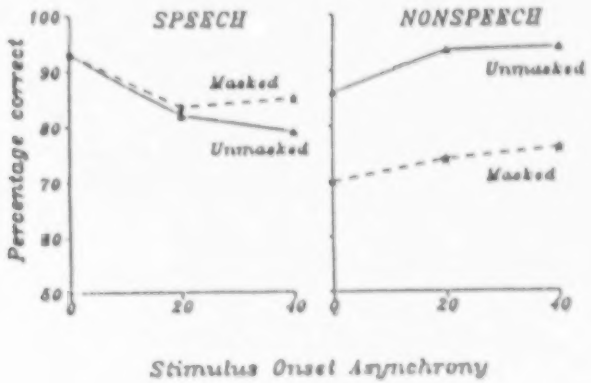


Figure 4. **The effect of backward mask intensity on the speech and nonspeech side of duplex percepts, compared with its effect on the perception of isolated transitions as chirps.**

Analysis of variance revealed significant main effects of perception mode, $F(2,20)=13.97$, $p<.0002$, and of mask intensity, $F(5,50)=9.00$, $p<.0001$, as well as a significant two-way interaction, $F(10,100)=3.13$, $p<.0016$. The interaction reflects the finding that perception of chirps in the duplex condition was more sensitive to mask intensity than perception of either speech or perception of chirps in the case of isolated formant transitions. However, this difference seemed to rest mainly on the decrease in performance with respect to the baseline condition. When the analysis was repeated with the baseline condition omitted, the two main effects remained reliable, but the interaction was no longer significant $F(8,80)=1.46$, $p>.18$. In a final analysis, only the speech and isolated chirp conditions were compared. No effect reached significance, although the main effect of mask intensity was close, $F(5,50)=2.27$, $p<.0618$.

The observation that masking has more of an effect on chirp perception in the duplex condition than on the perception of chirps heard in isolation would seem to pose a problem for a 'naive' masking accounts. We might avoid this problem by postulating an interaction between selective attention and the effects of the noise. Further discussion of this result is deferred until the completion of Experiment V, which asks about the generality of this result.

EXPERIMENT V

Experiment IV yielded the unexpected result that any differential effects of the ipsilateral mask on chirp vs. speech perception are peculiar to the duplex listening condition. To obtain further confirmation of a distinction between chirp perception in and out of the duplex condition, we conducted a final experiment which asked whether another type of acoustic manipulation—decreased amplitude—has a similar effect on the perception of isolated and duplex chirps. In designing this experiment we decided to improve upon our previous methodology by using duplex 'foils' in addition to the duplex stimuli (Repp & Bentin, 1984). In duplex foils, the 'base' is replaced by a full syllable which contains a third formant transition that is the opposite of the transition being presented to the other ear. The presence of such foils is designed to discourage listeners from using a correlation between speech and chirp percepts to improve their performance on either task.

The literature on speech perception in the duplex condition leads us to expect a high degree of tolerance of decreased amplitude. In his initial description of the duplex phenomenon, Rand (1974) reported that a 30 dB attenuation of isolated transitions of synthetic /ba/, /da/, and /ga/ did not impair correct labeling of the speech side of duplex percepts. Even with 50 dB attenuation of the transitions, labeling performance was still above chance (see also Cutting, 1976). However, no results have been reported regarding the effect of transition intensity attenuation on the chirp side of duplex percepts.

Methods

Subjects

The subjects were drawn from a pool of 25 Yale undergraduates who served as paid volunteers.

They included nine young men and women who had passed a selection criterion of better than 96% correct identification of the syllables and chirps which comprise the duplex stimuli (see below).

Stimuli

The /da/ and /ga/ syllables which formed the basis for the stimuli were similar to those employed in Experiment III and IV with the following differences. (1) The F3 transitions were 50 ms long. (2) The F3 transition for /ga/ began at 2018 Hz and rose to 2527 Hz. (3) The base contained a steady-state F3 at 2527 Hz. Six different intensity levels of the transitions were created by digital attenuation in 6 dB steps, resulting in levels of -12 to -42 dB relative to the base. The foils consisted of full /da/ and /ga/ syllables on one channel (the same as the base) paired with transitions cuing the opposite phonetic category on the other channel, with the different stimulus intensities occurring once each.

Procedure

To ensure a high level of performance subjects were screened for 96% or better speech labeling ability prior to the speech test, and for 96% or better chirp labeling ability prior to the chirp labeling tests. Each subject began the speech session by listening to a series of full /da/ and /ga/ syllables, presented to the left ear at the intensity they were to have in the duplex condition. Five alternations of the two syllables were presented first, followed by five blocks of 24 stimuli in random order. Subjects were instructed to write down "d" or "g" for each syllable, guessing if necessary; those who were confident of their performance were allowed to stop the pretest after two blocks. Subjects who made few or no errors on the first two blocks proceeded to the next condition, those who made an average of four or more errors per block were required to complete any remaining blocks. Of the subjects who completed all five blocks, those who averaged five or more errors per block were excused from further participation. Nine subjects had to be excused at this point. Two additional subjects passed the initial screening but were subsequently unable to label the full syllables and duplex syllables in the experiment proper; their data were also discarded.

The non-speech session began with a screening for the ability to distinguish the two chirps when the transitions were heard in isolation. The two transitions were first presented five times in alternation to illustrate the two categories of "high" and "low" (forcing subjects to use apparent pitch as a criterion) followed by five blocks of 24 transitions presented in random order to the right ear. The subjects were instructed to write down "H" or "L," guessing if necessary, and to complete at least two blocks. Criteria for passing subjects were the same as for the syllables, and four additional subjects had to be excused at this stage. One further subject's data were discarded because of random responding during the test phase.

In contrast to Experiment IV, the effect of transition intensity was tested with a random-presentation design. The stimuli were presented in five blocks of 48 stimuli, in which each block contained 36 true duplex stimuli (18 /da/ and 18 /ga/) and 12 foils (6 /da/ transitions presented with the full /ga/ syllable in the contralateral ear and 6 /ga/ transitions presented with the full /da/ syllable in the contralateral ear). The six intensity levels were used equally often with each stimulus type in each block. Thus, at each intensity level there were six duplex trials (three /da/ and three /ga/) and two foils (one /da/ and one /ga/). Across five blocks, there were 15 /da/ and 15 /ga/ transitions in duplex trials at each intensity.

The speech test session was presented immediately after the training and selection procedure. the subjects were required to listen to the ear receiving the base and to label the stimuli as beginning with "d", or "g," guessing as necessary. They were told to ignore any sounds that occurred in the other ear. The isolated chirp identification followed.

In the isolated chirp identification session, subjects listened to the same duplex tapes, but with the "base" channel disconnected. Because in this condition there could be no foils, subjects heard a total of 20 /da/ transition and 20 /ga/ transitions at each intensity. Since transitions at the lowest intensities were close to inaudible, a visual indication of each trial was presented to the subjects by a small light triggered by the onset of a stimulus on the left (disconnected) channel. Subjects were required to write 'H' or 'L' whenever the light flashed, whether or not they had heard anything. After completing identification of the isolated transitions, subjects listened to the duplex tape for a third time, with both output channels again connected to the earphones. The task was to identify the chirps in the right ear while ignoring the syllables in the left ear, guessing if necessary.

Results and Discussion

The accuracy of chirp and speech identification are graphed as a function of relative transition intensity in Figure 5. The left panel contains the

average percentage correct responses for speech identification, the right panel contains those for chirps. Chirp identification of isolated transitions is separated from that in the normal duplex stimuli, and in both panels, responses to duplex stimuli are separated from the responses to foil stimuli. For speech, the responses to foil stimuli have been graphed in terms of the accuracy with which subjects labeled the full syllable that replaced the base. For chirps, they have been graphed with respect to the accuracy with which subjects labeled the transition that was presented contralateral to the syllable.

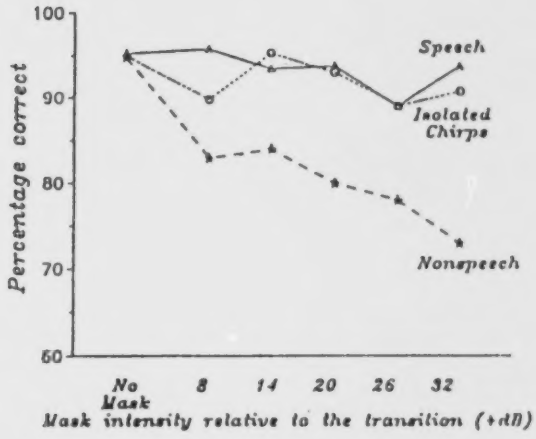


Figure 5. **The effect of transition amplitude on the speech and nonspeech side of duplex percepts, compared with its effect on the speech and non-speech side of duplex foils and on the perception of isolated transitions as chirps.**

Identification of the isolated chirps was quite good, and comparable to the identification of speech in the duplex condition. Apart from an unexplained decrease in the accuracy of chirp identification at the highest intensity, these two functions are practically identical, both reaching chance only at the lowest intensity value. In contrast, the identification of chirps in the duplex condition was adversely affected by the first step of attenuation. This pattern of results parallels that observed in Experiment IV, and was confirmed by an ANOVA which included all three conditions but omitted foil trials. That analysis revealed a main effect of condition (duplex speech vs. isolated chirp vs. duplex chirp), $F(2,16) = 12.61$, $p<.0005$, and an interaction with transition amplitude, $F(10,80) = 4.92$, $p<.0001$. Both effects were obviously due to the difference between the duplex chirp condition and the other two conditions. The main effect of transition amplitude was also significant. A separate analysis of variance, comparing identification of chirps in isolation to the identification of speech in duplex, revealed a significant main effect of transition attenuation, $F(5,40)=59.80$, $p<.001$, but no main effect of stimulus condition and no interaction.

Comparison of the results obtained with the duplex stimuli and those obtained with the foil stimuli offer insight into whether subjects were using speech percepts to bolster chirp identification, or vice-versa. The speech identification functions show that identification of the full syllables in the duplex foils is almost a mirror image of that of the true duplex stimuli. It is evident that, when the transition was at a very low intensity, the speech side of the foil stimuli was identified in accordance with the transition in the full-syllable "base" but that, as the transitions that were presented to the opposite ear were increased in amplitude, they competed with those in the full syllables and thereby lowered the accuracy of foil identification.

The non-speech functions in the right panel depart from this pattern in that both the function for the duplex chirp perception and the function for the foil chirp perception hover much closer to chance. Responses to chirps in the foil trials were at chance throughout, whereas responses in the duplex trials rose above chance only at the highest level of chirp intensity. This interaction approached significance in a separate test, $F(5,40) =2.36$, $p<.06$, but there was no significant main effect of transition amplitude for these two conditions. This lack of interaction makes it unlikely that subjects used the speech side of duplex percepts to improve their identification of the chirp side, for had this occurred, performance on the foil chirp identification trials would have been below chance at the lowest intensities. That is because the speech percepts in those cases were based on the full syllable, which contained the opposite transition of that presented to the other ear.

GENERAL DISCUSSION

This series of experiments addressed the possibility that, if speech and non-speech perception are dissociable modes of perception, it should be possible to show differential effects on each side of the duplex perception phenomenon. Since previous studies had successfully shown that phonetic manipulations affect the speech side

of duplex percepts but not the chirp side, we wondered whether it would be possible to find acoustic manipulations that would affect the chirp side but not the speech side. Our venture has been successful, for we have been able to show that both ipsilateral backward masking of the transition and decreasing transition intensity had more of a disruptive effect on the chirp side than on the speech side. In contrast, both a temporal asynchrony between base and transition and a mask that is ipsilateral to the base have marked effects on the speech side, but leave the chirp side unimpaired. Thus, we obtain evidence of a double dissociation which confirms the separability of the two forms of perception, and offers new support to previous suggestions that speech and nonspeech auditory stimuli are processed differently (Cutting & Pisoni, 1978; Liberman & Mattingly, 1985).

The result which remains to be discussed at this point is our somewhat unexpected finding that, whereas speech perception has been much the same when the base and the transition are dichotically presented in the duplex paradigm as when they are electronically fused and presented under normal binaural listening conditions (see Liberman et al., 1981; Mann & Liberman, 1983), we have obtained considerable evidence that chirp perception in the duplex condition is more vulnerable than the chirp perception in the case of isolated transitions. This result was anticipated by some data in Mann and Liberman (1983), which showed that the accuracy of chirp identification was considerably worse in the duplex condition than in the case of isolated transitions. It came to light in Experiment IV where we found that the effects of ipsilateral masking on chirp perception were markedly greater in the duplex condition than for isolated transitions. The greater vulnerability of chirp perception was also seen in Experiment V, where we found that a decrease in transition amplitude had a considerably greater effect on chirp perception in the duplex condition than on speech perception, and also a greater effect than on the perception of isolated chirps.

In seeking an explanation of why the duplex condition impairs chirp perception, there are several lines of reasoning to consider. We can discount a simple view that chirp perception is more vulnerable to the effects of stimulus degradation than is speech perception. Apparently, the results of Experiments II, IV and V, which showed that stimulus degradation penalized chirp perception in the duplex condition, support a view that speech perception is inherently more sensitive to marginal auditory information (possibly because the categories are better defined, more familiar, etc.). However, this account goes against the results of Experiment III, which showed that speech perception was penalized when the white noise occurred in the same ear as the base, whereas chirp perception was not. In addition, it goes against the results obtained with the isolated transitions in Experiments IV and V. Those results suggest that any greater susceptibility of chirp perception is somehow peculiar to the duplex condition.

Another potential account involves the concept of masking. In Experiment I we could argue that the base somehow masks the chirp. This could explain our result that, when presentation of the base is delayed, chirp perception is improved. It might further explain the finding that, in Experiment V, chirp perception in the duplex condition was more vulnerable than either duplex speech perception or isolated chirp perception to the effects of decreasing transition amplitude. Could masking also explain the results of Experiments II through IV? In considering a masking account, we would have to begin by postulating that the mask employed in our experiments was not a 'peripheral' one in the traditional sense, since a peripheral mask would be expected to influence all subsequent processing—speech and chirp alike. If a masking explanation is to succeed in these cases, differences in vulnerability to masking might, for example, reflect the greater length and complexity of the base vs. the isolated chirp. If length and complexity are the important factors, then one should obtain similar results when the base is replaced by some complex nonspeech sound. Note, however, that a masking account of this sort fails to acknowledge that the transition is perceived in two different manners, only one of which is 'masked.' This latter problem might be avoided by hypothesizing that speech perception of the integrated base and transition masks chirp perception of the isolated transition, but this would pose some conceptual problems. How can a stimulus mask itself?

We had previously mentioned the possibility that a 'stream-specific' mask might have operated in Experiment III. Let us now turn to considering a 'scene analysis' account (Bregman, 1978; 1981). Our results regarding the effects of the white-noise mask are compatible with an account in which masking is greatest when the white noise occurs in the same auditory 'stream' as the masked stimulus: The chirp is masked by a white noise which occurs in its 'stream' and the speech is

masked by a white noise which occurs in its 'stream.' If 'stream-specific' masking is the correct way to interpret the results of Experiment III, then we would predict that both speech and chirp perception would be 'released' from masking when the mask occurred in both ears simultaneously. That is because binaural presentation of the mask would cause the mask to be attributed to a separate stream.

However, the scene analysis account which we discussed above assumes that duplex perception results from the transition being incorporated into two separate 'streams,' one based on common location and one based on similar acoustic structure (i.e., common fundamental frequency contour). This would predict that dichotic presentation and similar acoustic structure are essential for the duplex phenomenon. Such a prediction is refuted by the demonstration of Whalen and Liberman (1987), who have shown that duplex perception is possible when neither constraint is met. In their study, listeners reported duplex percepts when a base and a sine-wave analog of a third-formant transition were presented to the same ear. There was neither a common F0 nor a spatial separation of chirp and transition, yet duplex percepts occurred when the transition was amplified to a level in excess of that which normally occurs.

We might also consider an account based on the differences in the ease of 'selective attention' to the speech and chirp percepts. A 'selective attention' account might go as follows: In the duplex phenomenon, listeners need to attend to one percept and ignore the other. Chirp perception is disrupted by the duplex condition because speech percepts are difficult to ignore, but the converse is not true, else the duplex condition should penalize both speech and nonspeech perception. Experiments I, II, IV and V are consistent with a view that the presence of a speech percept impairs the accuracy of chirp identification. In Experiment III, we could explain the observation that the presence of the ipsilateral mask impaired chirp perception but not speech perception, and the contralateral mask had the opposite effect by postulating that it is more difficult to attend to a given percept when a white noise is heard in the same spatial location. In the end, however, although selective attention may explain many of the present results, it leaves us with the basic question of why speech should be harder to ignore.

One might ascribe the fact that speech is hard to ignore to the fact that the speech percepts reflect the louder and longer portion of the stimuli, in which case one would predict that an equally 'complex' nonspeech sound would have the same effect on chirp perception, and this could, of course, be tested. One could also resort to the fact that speech categories are better defined, although our training and screening procedures attempted to control for this possibility, as evidenced by the high level of performance in identification of isolated chirps. One might even regard speech as 'harder to ignore' because it is mediated by a higher level of processing that has some privileged access to conscious introspection. The present results, for example, might be analogous to the word superiority effect where written 'words' are more tolerant of masking, etc. than individual letters because they involve a different level of processing (see Johnson, 1975 or 1977 for discussion of this effect). But the analogy between our results and the word superiority effect is questioned by one basic difference between the perception of letters in words and the perception of chirps in speech. Under normal reading conditions, when no mask is present and when stimulus duration is not severely limited, the letters which compose words are just as readily discernible as isolated letters. Yet under normal listening conditions, when speech signals are presented to the two ears at a reasonable intensity, listeners perceive formant transitions as part of the speech signal but fail to perceive them as chirps. It takes some rather bizarre manipulation of the stimulus—such as the duplex condition—to make a listener 'hear' the formant transition in speech signals as 'chirps.'

An account which we favor for its ability to explain this and other data in the speech perception literature is prompted by the view that, in the duplex listening condition, the presence of a speech percept takes precedence over chirp perception and serves to decreases the accuracy of chirp identification. This view has been set forth in several recent papers by Liberman and his colleagues (Liberman & Mattingly, 1987; Mattingly & Liberman, in press; Whalen & Liberman, 1987), and is based on a variety of evidence that the speech processing system (the phonetic module) has 'priority' over nonspeech systems. This priority accounts for the basic observation which began this paper, namely that, under normal listening conditions, formant transitions that are interpreted as support for one phonetic percept or another are not heard as nonspeech chirps. It further accounts for listeners' ability to hear transitions as 'chirps' under certain 'duplex' conditions: when they are amplified to some level in excess of that which normally occurs

(Whalen & Liberman, 1987), and when the the stimulus configuration presents certain discrepancies in fundamental frequency, spatial location, etc. Such conditions allow the listener to infer that two sources have contributed to the pattern (See Mattingly & Liberman, in press).

In the stimuli which this study employed, duplex perception of formant transitions was induced by separating the second or third formant transition from the remainder of the syllable and presenting it to the other ear, a maneuver which makes the chirps appears to 'float' away from the speech, and to have arisen from a separate spatial location (see Nygaard & Eimas, in press, for discussion of factors which influence the 'location' of the chirp percept). This artificial separation of transition and base leads listeners to perceive two different 'events'—a spoken syllable coming from one source, a chirp from another. However, there is a precedence of speech perception over nonspeech, such that the presence of the base supports a speech percept of the transitions and listeners behave as if the speech processor's interpretation of those transitions as part of a speech stimulus somehow interferes with their interpretation of the same transitions as nonspeech chirps. When the transitions are presented in isolation there is no speech percept, the precedence effect does not operate, and subjects achieve a higher level of performance.

Whalen and Liberman (1987) suggested that this interference exists because the speech processor operates first and subtracts energy from the signal, leaving only a residue for the nonspeech processor. However, other mechanisms (as yet unidentified) may cause this interference. What is important is that, with the 'precedence' account of the effect of the duplex condition on chirp perception, it seems reasonable enough that manipulations which alter chirp perception in the duplex condition need not have an effect on the perception of chirps which are heard in the case of isolated formant transitions. This was the result obtained in Experiments IV and V. In Experiment V, the manipulations of amplitude are a particularly appropriate test of the precedence hypothesis, for amplitude manipulations were the same means by which Whalen and Liberman (1987) were first able to demonstrate the precedence effect in monaurally-presented stimuli. Their experiment manipulated the relative amplitude of a sinusoidal third formant transition, and showed that when the amplitude of that transition was between 0 and -7 dB relative to the third formant of a synthetic speech 'base,' listeners heard both a 'whistle' (i.e., chirp) and the appropriate speech percept of "da" or "ga" As the amplitude of the transition was decreased, the nonspeech percept vanished, yet the speech percept continued to make use of transitions—even those that were 20 dB below the level of the base. Our stimuli are considerably different from theirs, yet we obtain a similar result: As can be seen in Figure 5, duplex chirp perception approaches chance as transition amplitude drops more than 6 dB below that of the original transition amplitude of -12 dB, whereas speech perception is still 70% accurate at as transition amplitude that is -18 dB below the original level.

The present data are consistent with Liberman and his colleagues' hypothesis about the (biological) 'specialness' of the speech processor and its priority over other forms of perception. They can even offer some insight into the conditions under which that priority is effective, if we consider the results of Experiments I, II and III more carefully. In those experiments it appears that chirp identification is most accurate when the base and transition are asynchronous, that is, when integration of the base and transition is disrupted. There is also some indication that the effects of an ipsilateral mask are reduced when asynchrony is present, though this result is less clear. Temporal asynchrony does not prevent the listener from hearing speech, it merely disrupts the likelihood that the transition will be incorporated into the speech percept. Apparently the mere presence of speech perception is not as critical as is competition for the same acoustical information. Hence we suggest that speech perception takes precedence when the speech and non-speech modes are 'competing' for interpretation of one and the same transition.

REFERENCES

Bregman, A. S. (1981) Asking about the 'What for' question in auditory perception. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Bregman, A. S. (1978) The formation of auditory streams. In Jean Requin (Ed.) Attention and performance VII. Hillsdale, NJ: Lawrence Erlbaum Associates.

Cutting, J. E. (1976). Auditory and linguistic processes in speech perception: Inferences from six fusions in dichotic listening. *Psychological Review, 83*, 114-140.

Cutting, J. E., & Pisoni, D. B. (1978). An information processing approach to speech perception. In J. F. Kavanagh & W. Strange (Eds.), *Speech and language in the laboratory school and clinic* (pp. 38-72). MIT Press, Cambridge, MA.

Johnson, N. F. (1977) A pattern unit model of word identification. In D. LaBerge & S. J. Samuels (Eds.), *Basic processes in reading: perception and comprehension*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Johnson, N. F. (1975) On the function of letters in word identification: Some data and a preliminary mode. *Journal of Verbal Learning and Verbal Behavior, 14*, 17-29.

Liberman, A. M. (1979). Duplex perception and integration of cues: Evidence that speech is different from nonspeech and similar to language. In E. Fischer-Jrgensen, J. Rischel, & N. Thorsen (Eds), *Proceedings of the Ninth International Congress of Phonetic Sciences* (Vol 2, pp. 468-473). Copenhagen: University of Copenhagen Press.

Liberman, A. M., Isenberg, D., & Rakerd, B. (1981). Duplex perception of cues for stop consonants: Evidence for phonetic mode. *Perception and Psychophysics, 30*, 133-143.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21*, 1-36.

Mann, V. A. (1986). Distinguishing universal and language specific levels of speech perception: Evidence from Japanese perception of /l/ and /r/. *Cognition, 15*, 169-196.

Mann, V. A., & Liberman, A. M. (1983). Some differences between phonetic and auditory modes of perception. *Cognition, 14*, 211-235.

Massaro, D. W. (1970). Pre-perceptual auditory images. *Journal of Experimental Psychology, 85*, 411-417.

Mattingly, I. G., & Liberman, A. M. (in press). Speech and other auditory modules. In G. M. Edelman, W. E. Gall, & W. M. Cowan (Eds.) *Signal and sense: Local and global order in perceptual maps*. New York: Wiley.

Mattingly, I. G., Liberman, A. M. Syrdal, A. M., & Halwes, T. (1971). Discrimination in speech and nonspeech modes. *Cognitive Psychology, 2*. 131-157.

Nygaard, L. C., & Eimas, P. D. (in press) A new version of duplex perception: Evidence for phonetic and nonphonetic fusion. *Journal of the Acoustical Society of America*.

Nusbaum, H. C., Schwab, E. C., & Sawusch, J. R. (1983) The role of "chirp" identification in duplex perception. *Perception & Psychophysics, 33*, 469-474.

Pilch, H. (1979). Auditory phonetics. *Word, 29*, 148-160.

Rand, T. C. (1974). Dichotic release from masking for speech. *Journal of the Acoustical Society of America, 55*, 678-680.

Repp, B. H. (1981). Two strategies in fricative discrimination. *Perception & Psychophysics, 30*, 217-227.

Repp, B. H. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin, 92*, 81-110.

Repp, B. H. & Bentin, S. (1984). Parameters of spectral/temporal fusion in speech perception. *Perception & Psychophysics, 36*, 523-530.

Repp, B. H., Milburn, C., & Ashkenas, J. (1983) Duplex perception: confirmation of fusion. *Perception & Psychophysics, 33*, 333-337.

Whalen, D., & Liberman, A. M. (1987). Speech perception takes precedence over nonspeech. Science, 237, 169-171.

FOOTNOTES

*Appears in *Journal of the Acoustical Society of America, 88(1)*, 64-74.

†Also Department of Psychology, The Hebrew University, Jerusalem, Israel.

††Also Department of Cognitive Sciences, University of California, Irvine.

Haskins Laboratories Status Report on Speech Research
1991, SR-105/106, 187-214

The Influence of Spectral Prominence on Perceived Vowel Quality*

Patrice Speeter Beddor† and Sarah Hawkins††

Research indicates that when the first and second formants of a vowel are separated by less than about 3.5 Bark, perception of its height and some other aspects of its quality is determined by some weighted average of the low-frequency spectrum, rather than by particular harmonic or hypothetical formant frequencies (as is the case with more widely spaced formants). This spectral averaging has been called the center of gravity (COG) effect. Although the existence of the effect is generally accepted, the factors that govern it are poorly understood. One possibility is that the influence of the spectral envelope on perceived vowel quality increases as low-frequency spectral prominences become less well defined. A series of three experiments examined this possibility in (1) nasal vowels, where the lowest spectral prominence is broader and flatter than that of oral vowels, (2) one- versus two-formant vowels with bandwidths appropriate for oral vowels, and (3) two-formant vowels with very narrow or very wide bandwidths. The results of these experiments show that when two or more spectral peaks lie within 3.5 Bark of one another, F1 and the centroid (an amplitude-weighted average frequency that estimates the COG in the low-frequency spectrum) roughly determine the boundaries within which the perceptual COG lies; the frequencies of spectral peaks dominate responses when formant bandwidths are narrow, whereas overall spectral shape exerts more influence when spectral prominences are wide. Assuming that all vowels undergo the same processing, we suggest that vowel quality, particularly height, is determined both by the frequency of the most prominent harmonics in the low-frequency region and by the slopes of the skirts in the vicinity of these harmonics. These two effects are most clearly separable in vowels with poorly-defined spectral prominences whose shape cannot be adequately described by specifying the frequencies and degree of prominence of just one or two harmonics, or hypothetical formant peaks.

INTRODUCTION

It is generally recognized that perception of vowel quality depends on the relative frequencies of the first two or three formants. Experiments indicate, however, that whenever two adjacent spectral peaks are close in frequency, a similar vowel quality can be achieved by substituting a single peak whose center frequency falls between the center frequencies of the original peaks, and is often dependent on the relative intensities of the two original peaks. This phenomenon, sometimes called formant averaging, or, more generally, spectral averaging or spectral integration, was first demonstrated by Delattre, Liberman, Cooper, and Gerstman (1952) for F1 and F2 of back vowels, which are relatively close in frequency. Subsequent experiments demonstrated spectral averaging for F1 and F2 (Assmann, 1985; Miller, 1953), for higher formants (Bladon & Fant, 1978; Carlson, Fant, & Granström, 1975; Carlson, Granström & Fant, 1970; Miller, 1953), and for the first harmonic and F1 (Carlson et al., 1975; Fujisaki & Kawashima, 1968; Traunmüller, 1981).

In an extensive series of experiments in which listeners matched, for vowel quality, one-formant

This research was supported by N.I.H. Grants NS-07237 and RR-05596 to Haskins Laboratories, and N.I.H. Postdoctoral Grant NS-07196 to Patrice Beddor. In the course of this work we received help from many colleagues, and we would particularly like to thank Brian Glasberg, Dennis Klatt, Björn Lindblom, Richard McGowan, Brian Moore, Roy Patterson, Philip Rubin, and Ann Syrdal.

with two-formant synthetic vowels, Chistovich and her colleagues investigated parameters that affect spectral integration in the F1-F2 region (Bedrov, Chistovich, & Sheikin, 1978; Chistovich, 1985; Chistovich & Lublinskaya, 1979; Chistovich, Sheikin, & Lublinskaya, 1979). They concluded that spectral integration would occur when the center frequencies of adjacent spectral peaks were within 3 to 3.5 Bark of one another, but not when the peaks in question spanned a greater frequency range. When the two spectral peaks of the reference vowel fell within this critical range but were of different amplitudes, then the center frequency of the best-match one-formant stimulus shifted towards the higher-amplitude formant. Chistovich et al. termed this inducible shift in phonetic quality of vowel approximations the "center of gravity" (COG) effect.

The Leningrad group's studies suggested that the COG can be influenced by both the overall spectral shape and the location and envelope of the major spectral peaks. They hypothesized that the centroid, an amplitude-weighted measure of mean frequency in the critical frequency range, would describe the effective perceptual center of gravity. Investigations of the details of the COG effect, however, indicate that Chistovich's formulation of the concept in terms of frequency and amplitude must be modified in a number of ways. For example, by trading amplitude against frequency such that there would be no change in the perceptual COG if formant frequency and amplitude were equally important, Klatt (1985) demonstrated that formant frequency is a stronger determinant of the COG than is formant amplitude. These data supported Ainsworth and Millar's (1972) earlier observation of the lack of influence of formant amplitude in determining vowel quality. Klatt and Ainsworth and Millar argued that vowel quality is mediated by a peak-picking mechanism rather than by a mechanism that integrates spectral energy within certain frequency bands; other investigators have made similar points (e.g., Carlson et al., 1975; Carlson & Granström, 1979; Paliwal, Ainsworth, & Lindsay, 1983).

Assmann (1985) suggested that formant amplitude may affect the perceptual COG when the experimental matching stimuli have only a single formant, as in most of Chistovich et al.'s experiments. Formant frequency apparently predominates when multiformant stimuli are used in other paradigms, for instance Ainsworth and Millar's identification tasks or Klatt's similarity judgments. Assmann's (1985) own experiments indicate that formant amplitude can have small effects on the COG under some circumstances, but his data are incompatible with the hypothesis of a general mechanism for averaging formant frequency and amplitude. When six-formant synthetic stimuli with a constant 250 Hz separation between F1 and F2 (less than 3 Bark) were matched with similar reference stimuli in which the amplitude of F2 was systematically varied, the COG shifted when the amplitude of F2 was raised, but not when it was lowered.

In a second experiment, Assmann (1985) found that degree of spectral integration was influenced by the absolute value of certain parameters. Vowel identification was affected by shifts in formant amplitude only when the stimuli bore similarities to a woman's voice (F0 of 250 Hz instead of 125 Hz, and an increase in the frequency of the higher formants). Moreover, COG shifts in this experiment were greatest when the separation between the first two formants exceeded 3.5 Bark.

These and other data suggest that the perception of vowel quality does not depend exclusively upon formant frequency, but that we do not understand the conditions under which other parameters have an effect. Although a number of models of the perception of vowel quality have been offered (e.g., Carlson et al., 1975; Chistovich et al., 1979; Paliwal et al., 1983), it is probable that none as yet takes into account all the relevant parameters. It is particularly important to evaluate the effect of these parameters in sounds resembling natural speech. As we have noted, responses to one- and two-formant stimuli can be very different from responses to more fully specified vowels, and this difference may be crucial since there is some evidence that formant averaging may be specific to speech-like sounds (Delattre et al., 1952; Traunmüller, 1982). On the other hand, manipulation of the low-frequency spectrum of multiformant vowels usually affects higher frequencies as well, so it is difficult to synthesize multiformant vowels that differ systematically in only one part of the spectrum. Experiments using the more natural multi-formant and the better-controlled one- or two-formant stimuli both have their place in investigating vowel quality.

The class of nasal vowels offers the opportunity to assess for multiformant stimuli whether it is the first formant peak in particular that is influential in determining perceived vowel quality, or whether it is the low-frequency spectrum in general. The principal spectral property that distinguishes nasal from nonnasal vowels is the

amplitude and bandwidth of the lowest spectral prominence (Delattre, 1954; Hawkins & Stevens, 1985; House & Stevens, 1956). Specifically, when the velum opens during production of a nasalized vowel, the nasal cavity contributes pole-zero pairs which modify the spectrum of the nonnasal vowel. For most vowels, the greatest effect is in the low frequencies. The lowest nasal pole-zero pair interacts with the lowest resonance of the oral cavity such that the spectrum output at the lips is distinctly different at low frequencies from that for the corresponding nonnasal vowel (Fant, 1960; Fujimura & Lindqvist, 1971; Stevens, Fant, & Hawkins, 1987). The detailed spectral shape of nasal vowels at low frequencies varies with the frequency of the lowest resonance due to the oral cavity, but as a rule the lowest spectral prominence of a nasal vowel has a lower amplitude and wider bandwidth than that of the corresponding nonnasal vowel. Measurements of naturally-spoken nasal vowels from a number of languages indicate that the (shifted) F1 and lowest-frequency nasal formant are usually within 3.5 Bark of one another (Beddor, 1982).

When the spectrum has a broad region of relatively undifferentiated peaks and valleys in the first formant region, as with nasal vowels, the overall spectral envelope may be more influential in determining vowel quality than when spectral peaks are well defined, as they are in oral vowels. There is substantial phonological evidence that, historically, nasalization of a vowel results in a shift in that vowel's quality, especially in the height domain (Beddor, 1982; Bhat, 1975; Ohala, 1974; Ruhlen, 1978; Schourup, 1973). In general, these shifts are such that the height of high and mid nasal vowels tends to fall and that of low nasal vowels tends to rise; the direction of these shifts is exactly what would be predicted if the low-frequency nasal formant added by nasal coupling had a strong effect on the perceptual COG of nasal vowels.

Other data are also consistent with the suggestion that the influence of the spectral envelope on perceived vowel quality increases as low-frequency spectral prominences become less well defined. Assmann (1985), for example, found a stronger COG effect for vowels with a high fundamental frequency; presumably the effective formant bandwidths were wider when the fundamental was high. This effect is predicted by auditory masking studies and by physiological studies of the auditory nerve which show that neurons over a wide range of characteristic frequencies respond in synchrony with the frequency of a narrow-bandwidth, high-amplitude spectral peak, but respond at their characteristic frequencies when the stimulating spectral prominence has a wide bandwidth of sufficient amplitude (Sachs and Young, 1980). In the latter case, the influence of the first formant frequency may be expected to be reduced. Consequently, models of perception of vowel quality which emphasize peak frequency (e.g., Carlson & Granström, 1979) or the amplitude of a small number of the most prominent harmonics (Assmann & Nearey, 1987; Carlson et al., 1975; Mushnikov & Chistovich, 1972) may be less appropriate for nasal than for nonnasal vowels, and indeed for all vowels that lack sharply defined formants.

This paper reports three experiments which together examine some aspects of the influence of spectral prominence as a determinant of vowel quality, especially vowel height. The first two experiments examined the COG effect in multiformant nasal (compared with nonnasal) vowels, and in two-formant (compared with one-formant) nonnasal back vowels, a design similar to that originally used by Chistovich et al. (1979). In a third experiment, the influence of spectral prominence, and specifically formant bandwidth, was assessed by comparing two-formant vowels that had moderate bandwidths with stimuli that had two formants of either very narrow or very wide bandwidth.

I. EXPERIMENT 1: MULTIFORMANT NASAL AND NONNASAL VOWELS

In this experiment, which was designed to test the center-of-gravity effect in natural-sounding nasal vowels, multiformant nasal vowels were compared with multiformant oral vowels. The oral-nasal difference was achieved by manipulating low-frequency spectral characteristics in such a way that the manipulations had little effect on the higher frequencies.

A. Method

1. Stimuli

Five sets of synthetic nasal and oral vowels were generated on the Haskins serial formant synthesizer written by Mattingly. Each 360-ms stimulus consisted of steady-state vowel formants, with fundamental frequency and amplitude decreasing over the final 120 ms.

The five nasal vowel stimuli, [ĩ ɛ̃ æ̃ ɑ̃ ɔ̃] were each synthesized from five poles and an additional pole-zero pair in the vicinity of the first (oral) pole, F1. The spectral characteristics of the synthetic

nasal vowels were based on acoustic analyses of natural tokens from several languages (Beddor, 1982). Parameter values for the frequencies and bandwidths of the poles and the zero are given in Table 1. Figure 1 shows 512-point DFT spectra of the synthesized nasal vowels. Consistent with acoustic theory (Fant, 1960; Fujimura & Lindqvist, 1971), the frequency of the peak labeled FN, the nasal formant, was higher than the peak labeled F1 in the high and mid vowels, but less than F1 in the low vowels. Thus the effect of the nasal formant is to raise the COG in high and mid nasal vowels relative to their oral counterparts, but to lower the COG in low vowels.

Table 1. *Nasal vowels. Numbers represent the frequency (in Hz) of the poles and zero (FZ) for the five nasal vowel stimuli. F4 was 3500 Hz and F5 was 4500 Hz. Bandwidths were 80, 120, 175, 300, and 300 Hz for F1-F5 respectively, and 200 Hz for both FN and FZ.*

Vowel	F1	F2	F3	FN	FZ
[ĩ]	295	2275	2900	1000	1250
[ẽ]	450	2000	2550	700	560
[æ̃]	640	1900	2500	350	540
[ɑ̃]	725	1150	2450	400	625
[õ]	450	775	2300	664	540

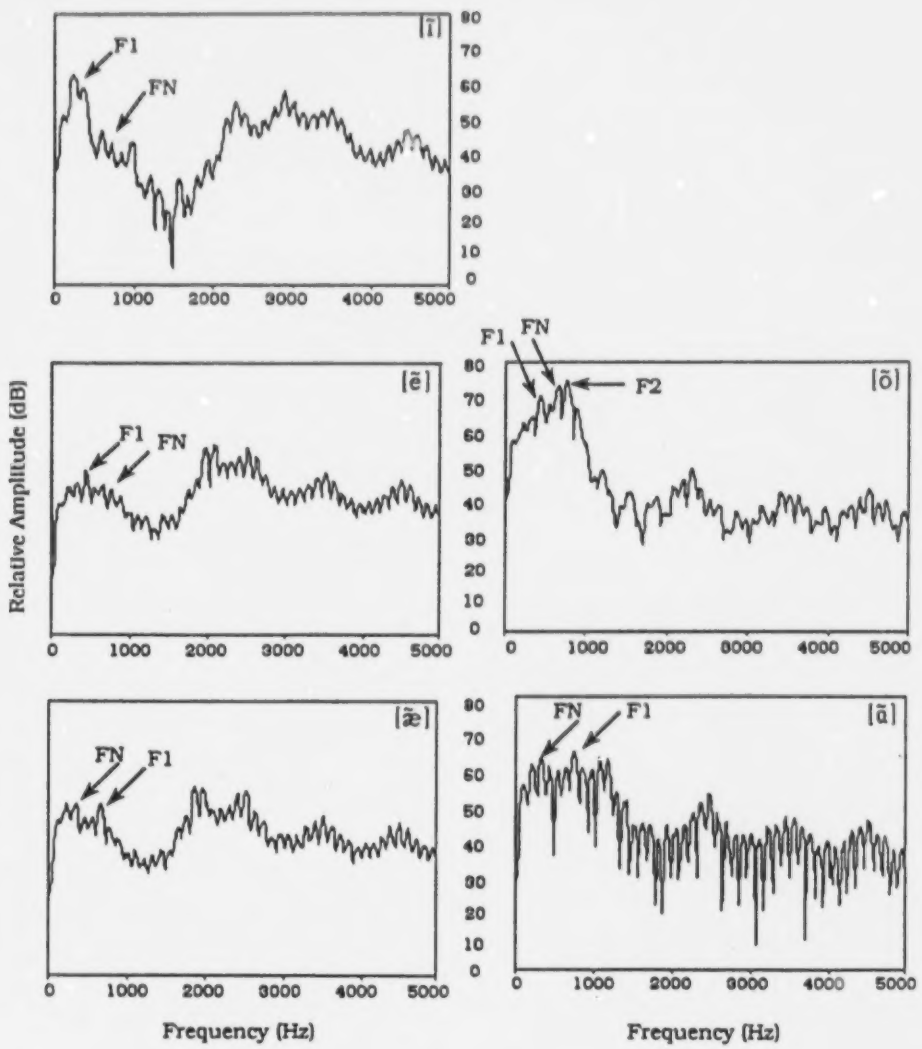


Figure 1. DFT spectra of the five synthetic nasal vowel stimuli of Experiment 1.

For each nasal vowel, a series of corresponding oral vowels was synthesized by omitting the nasal pole-zero pair and systematically varying the frequency of F1. The perceptual effect of manipulating F1 was to change the height of the vowel. Figure 2 shows F1 values for the members of the five oral vowel series, *i*, *e*, *a*, *æ*, *o*; each dot represents the F1 frequency of a single oral vowel stimulus. (See Table 2 for parameter values of these oral stimuli.) The change in frequency of F1 between successive stimuli in a series was approximately 10% of the average F1 frequency for that series, meaning that differences were larger for lower vowels (70 Hz for *a*,, 60 Hz for *æ*, 45 Hz for *e*, 40 Hz for *o*, and 32 Hz for *i*). Each series included two vowels of special interest: an *F1 match*, where F1 frequency of the oral vowel was identical to that of the corresponding nasal vowel (indicated by the solid horizontal line in Figure 2); and a *centroid match*, where the centroid of the oral vowel was the same as that of the corresponding nasal vowel (indicated by the dashed line).

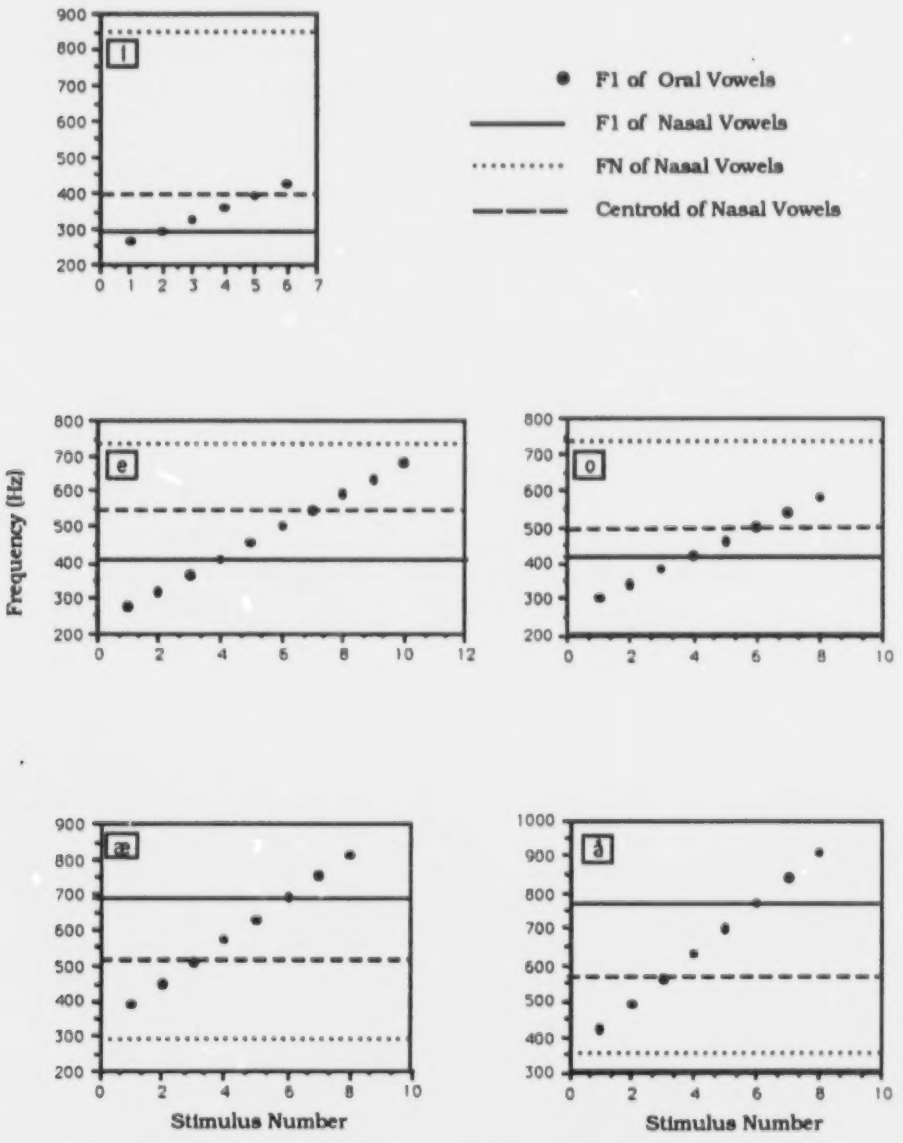


Figure 2. **F1 frequencies (shown by solid dots) for the stimuli comprising the oral series for each vowel set of Experiment 1. F1 frequency (solid line), FN frequency (dotted line), and centroid frequency (dashed line) of the corresponding nasal vowel are shown for reference. (See text for explanation of centroid measure.)**

These frequency matches were based on autoregressive LPC spectra, but were corroborated to within 5 Hz with measures from 512-point DFT spectra.

Table 2. *Oral vowels. Numbers represent the frequency (in Hz) of F1 for the members of the five oral vowel series. F2-F5 values and all bandwidth values are as specified for the corresponding nasal vowels in Table 1.*

	Stimulus number									
	1	2	3	4	5	6	7	8	9	10
i	263	295	327	359	391	423	-	-	-	-
e	275	320	365	410	455	500	545	590	635	680
æ	390	450	510	570	630	690	750	810	-	-
a	420	490	560	630	700	770	840	910	-	-
o	300	340	380	420	460	500	540	580	-	-

2. Centroid measurements

The centroid, a weighted average frequency, is a measure of center of gravity. It is computed from the mean frequency of the area under the spectral curve within specified frequency and amplitude ranges according to the formula

$$X_{CEN} = \frac{\sum_{i=1}^{n} (X_i Y_i)}{\sum_{i=1}^{n} Y_i}$$

where X = frequency (Hz) and Y = log magnitude (dB). This formula is the same as that published by Chistovich (1985) except that she used a Bark frequency scale whereas we used a linear one. The influence of type of frequency measure on our data is addressed in the General Discussion below.

The centroid of a particular stimulus may of course vary substantially depending on the frequency and amplitude ranges specified in the centroid calculation. However, changes in these ranges do not always have the same effect on all types of stimuli. For example, the location of the upper frequency cutoff has a substantial effect on the centroids of the multiformant vowels of this first experiment but only a negligible effect for the one- and two-formant vowels of our later experiments, due to the presence of higher frequency information in the former and its absence in the latter. Consequently, a single criterion for frequency and amplitude ranges cannot be satisfactorily applied to a broad range of stimuli.

This lack of generality of the centroid measure raises a number of issues. Our interpretation is that the centroid is at best a purely descriptive measure of perceived vowel quality. Our purpose was to assess the extent to which the centroid as used by Chistovich and her colleagues could be usefully applied to stimuli whose formants varied in spectral prominence. We decided, then, to use centroid measures that captured as closely as possible the spirit of Chistovich's formulation, and at the same time had some psychoacoustic validity. Our approach was to use an area of the spectrum that included all the formant peaks in which we were interested, and extended far enough down from the peaks to include as much of their skirts as the auditory system appeared likely to process (cf. Carlson, Granström, & Klatt, 1979). In deciding upon the criteria used for each particular set of stimuli, a number of values were tried; the chosen ones satisfied the above criteria and were relatively insensitive to small changes in the spectral envelope in the regions of least amplitude.

The frequency range applied to the oral and nasal vowels of this experiment was 100-1100 Hz, except that, for the *a* series, the upper limit was extended to 1400 Hz. These ranges were selected so as to include all low-frequency formant peaks and skirts: F1 in all vowels, and also FN in the nasal vowels, and F2 in the back vowels. The lower amplitude limit was taken as the lowest-amplitude point of the spectral envelope in the 100-1100 (or 1400) Hz region; hence this limit was different for each stimulus. Figure 3 illustrates the centroid measure as applied to the vowels [ɛ̃] and [æ̃]. The thick vertical lines delimit the frequency range of 100-1100 Hz and the connecting horizontal line sets the lower amplitude limit. The center frequency or centroid of this area is shown by the dashed vertical line.

Figure 4 illustrates, for the *e* series, F1 and centroid matches between oral and nasal stimuli. For clarity, LPC rather than DFT spectra are shown. The upper panel shows spectra of the nasal vowel (solid line) and the F1 match from the oral series (dashed line). The frequency of F1 is the same in both stimuli. The lower panel shows spectra of the nasal vowel (solid line) and the centroid match of the oral vowel series (dashed line). Here, F1 of the oral vowel falls between the two low-frequency peaks, F1 and FN, of the nasal vowel; although these two spectra share no peak frequency below 1100 Hz, they have the same centroid frequency in this region.

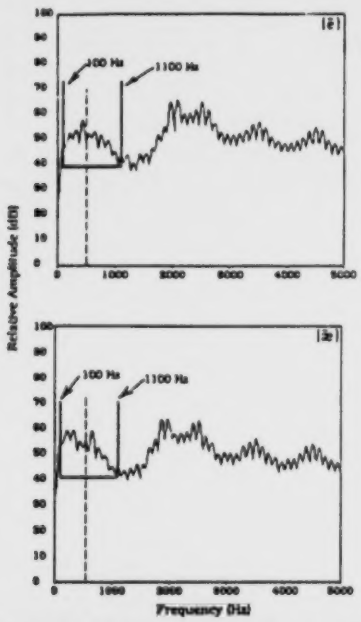


Figure 3. **Illustration of the centroid measure applied to two of the nasal vowel stimuli of Experiment 1, [ɛ̃] (top panel) and [æ̃] (bottom panel). The heavy vertical lines at 100 and 1100 Hz delimit the frequency range, and the connecting horizontal line sets the lower amplitude limit. The dashed vertical line indicates the centroid of this area.**

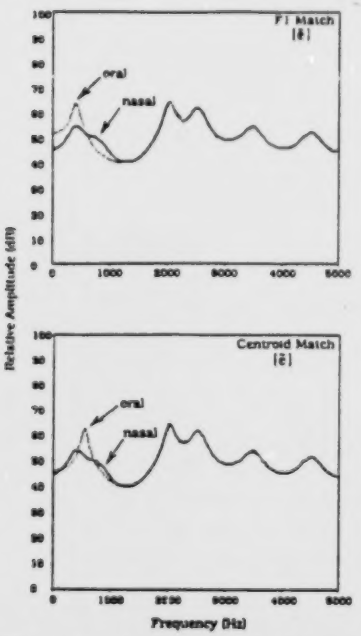


Figure 4. **LPC spectra of the F1-matched (top panel) and centroid-matched (bottom panel) stimulus pairs for vowel set *e* of Experiment 1. Oral (dashed line) and nasal (solid line) vowels are identical in F1 frequency in the top panel and are identical in centroid frequency in the bottom panel.**

3. Subjects

Twenty paid student volunteers, aged 18 to 25 years, participated in the experiment. All were native speakers of American English with no known hearing loss and no training in phonetics.

4. Procedure

In the test sequences for the five vowel sets, every nasal vowel was paired with each oral vowel from the appropriate series (i.e., for each vowel set, one nasal vowel was paired with 6-10 oral vowels), with the nasal reference vowel always being the second member of the pair. For each vowel set, these oral-nasal pairs were arranged in two ordered sequences: ascending sequences (from the lowest frequency of F1 in the oral vowels to the highest) and descending sequences (from the highest frequency to the lowest). A pilot study in which listeners selected the "best-match" oral-nasal pair from these sequences showed that matches tended to fall in the middle of each vowel set. To control for possible range effects, three truncated ordered sequences were derived from the complete series for each vowel by omitting 1 or 2 members from the beginning or end of each complete series. The three truncated versions of each of the five vowel sets were then arranged in random order, for a total of 15 trials. The inter-stimulus interval between members of an oral-nasal pair was 0.5 s and the interval between pairs in the ordered sequences was 1 s; listeners controlled all other time intervals.

Before testing, subjects were given a brief description of the stimuli. They were told that each trial would consist of several vowel pairs, that the first member of each pair varied across the series while the second member stayed the same, and that these pair members were "oral vowels" and "nasal vowels" respectively. It was explained that nasal vowels usually occur in English in the context of *m* or *n*, e.g., *mom* (versus the oral vowel in *Bob*), *man* (versus *bad*), and *moan* (versus *boat*).

Individual listeners were tested on-line in a sound-attenuated booth. Stimuli were presented binaurally over TDH-39 earphones. At the onset of each trial, ascending and descending truncated sequences were presented for that particular vowel series. (The relative order of ascending and descending sequences was counter-balanced across trials.) Listeners could then request, by keyboard commands, repetitions of either of the sequences or of individual oral-nasal pairs from the sequence. For each trial, listeners were instructed to select that pair in which the oral vowel was the most similar to the nasal standard;

this "best-match" pair was circled on a printed score sheet. Listeners were encouraged to listen to the sequences and to individual pairs as many times as needed to feel confident about the best-match decision. Average testing time was approximately 45 minutes.

B. Results

We hypothesized that if perceived nasal vowel height were determined by the spectral COG, then the closest perceptual oral-nasal match would be the centroid-matched pair. If perceptual integration of F1 and FN did not occur, then we might expect the F1-matched vowels to be perceptually more similar. Listeners' responses suggest that, except for [ĩ], reality lies somewhere between these two extreme possibilities. Figure 5 gives the responses, pooled across the 20 subjects, to the five vowel sets. These functions show that oral vowels judged most similar to non-high nasal reference vowels had F1 (and centroid) frequencies that fell between the centroid and F1 frequencies of the nasal reference. The F1-matched pair only accounted for between 2% and 12% of the responses to [ɛ̃ æ̃ ɑ̃ ɔ̃], whereas it accounted for over 70% of the responses to the high vowel [ĩ].

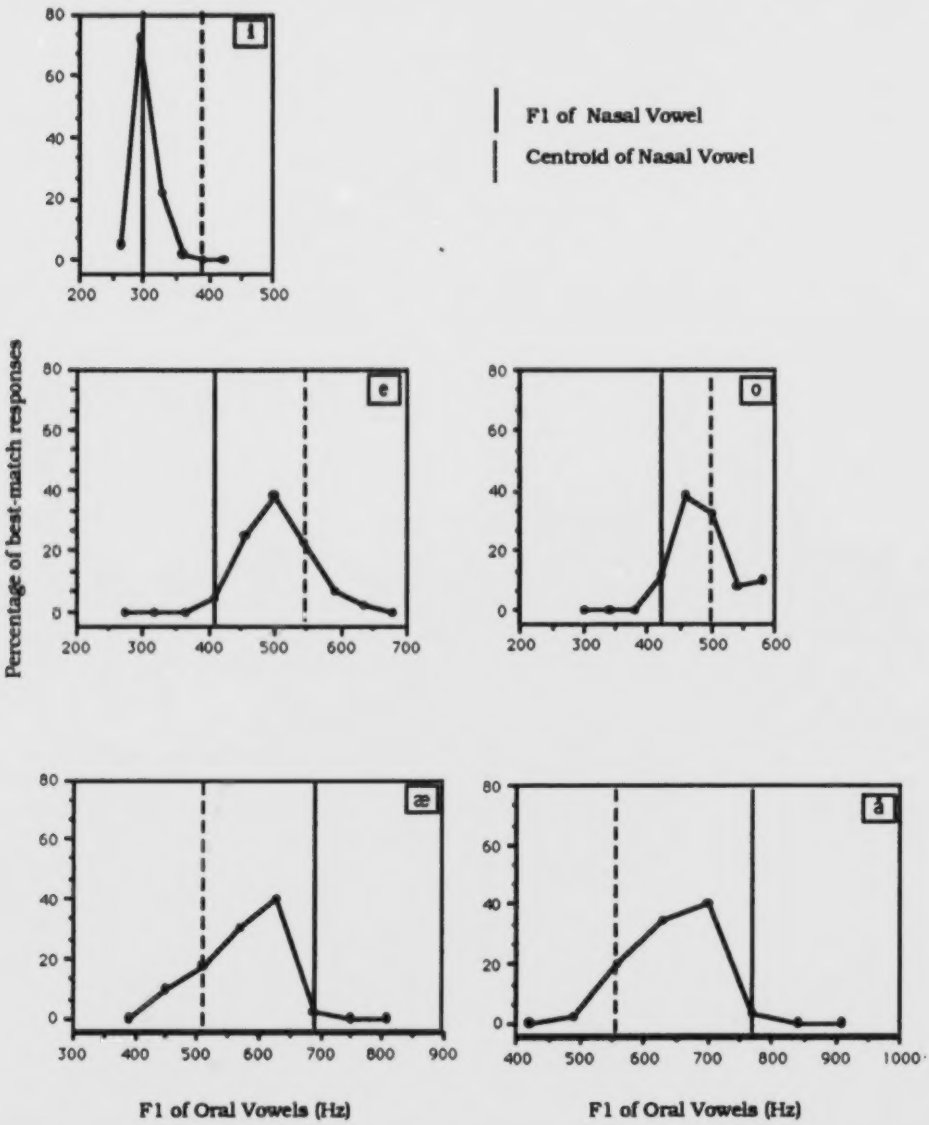


Figure 5. Percent "best-match" responses to the oral-nasal pairs for each vowel set of Experiment 1. F1 and centroid frequencies of each nasal vowel are shown as vertical lines.

In each case, responses were skewed towards the centroid frequency of the nasal reference vowel, and this skew meant that responses were significantly closer to the centroid than to F1 in the four non-high vowel sets. (For this calculation, each subject's mean response was subtracted from the stimulus number of the centroid match and of the F1 match of a given vowel set, and t-tests were performed on the two resultant distributions. For *e, æ, o*, $t(19) = 3.27, 3.41, 4.87$ respectively, $p < .01$; for *a*, $t(19) = 2.24$, $p < .05$.) In the high vowel set *i*, on the other hand, listeners' responses were significantly closer to the frequency of F1 ($t(19) = 18.5$, $p < .01$), as was predicted since the difference between F1 and FN frequencies exceeded 3.5 Bark for this vowel.

Even for *i*, however, the function is sufficiently skewed that the mean stimulus number of subjects' responses differed significantly from that of the F1 match ($t(19) = 2.68$, $p < .05$). For the other four vowels, as Figure 5 suggests, the differences between listeners' mean responses and the F1-match stimulus number were more strongly significant, all at $p < .01$ or better ($t(19) = 11.87$ for *e*, 15.88 for *æ*, 14.45 for *a*, and 10.97 for *o*).

The skew towards the centroid value of the nasal reference means that there were more choices of mid oral vowels (*e, o*) with higher F1 frequencies than in the nasal reference, and more choices of low oral vowels (*æ, a*) with lower F1 frequencies than in the nasal reference. That is, the skews were in opposite directions for mid and low vowel sets.

C. Discussion

The results of this experiment substantiate previous speculations on the perception of height in nasal vowels. Joos (1948), for example, suggested that French /ɛ̃/ sounded like [æ̃] because the average frequency of F1 and FN in nasal /ɛ̃/ corresponded to F1 in oral /æ/. Similarly, Fant (1960), Beddor (1982), and Wright (1986) hypothesized that shifts in perceived vowel height accompanying nasal coupling might be due to the additional low-frequency nasal resonance. Empirical support for these speculations is provided by the present findings: listeners' mean responses always fell between the frequency of FN and F1, which meant that the effect of nasalization was to raise the perceived height of low nasal vowels (where nominal FN frequency is less than nominal F1 frequency) and lower the perceived height of high and mid nasal vowels (where nominal FN frequency is greater than nominal F1 frequency). These data also provide a perceptual explanation for the linguistically widespread pattern of sound changes, noted in the Introduction, whereby vowel nasalization tends to have a centralizing effect on vowel height (see Beddor, Krakow, & Goldstein, 1986, for further discussion).

While the data of this experiment also confirm the basic principles of the COG model, they raise questions about its detailed predictions. The model is confirmed in that F1 appears to dominate the perception of vowel height only when adjacent spectral peaks are sufficiently far apart. The present experiment was not designed to test the validity of the critical cutoff of 3 - 3.5 Bark, which has in any case been attested by others (e.g., Bladon, 1983; Syrdal, 1985; Syrdal & Gopal, 1986). But responses to the vowel [ĩ], whose F1-FN spacing was 4.5 Bark, are consistent with this critical value: these responses were quite different from the responses to the other four vowels, all of which had F1-FN spacings of less than 3.5 Bark.

The data give little support for the validity of the centroid measure we used as a measure of the perceptual center of gravity in these nasal vowels. For nasal vowels whose F1 and FN are within 3.5 Bark of one another (i.e., for all except [ĩ]), the response curves of Figure 5 are broadly within the frequency range encompassed by F1 and the centroid, indicating a wide range of first-formant frequencies which listeners judge as reasonably acceptable matches. Rather than marking the perceptual COG, then, the centroid frequency as we calculated it appears to provide one *boundary* to a range of frequencies within which the perceptual COG lies; the other boundary to this range of frequencies is the first formant.

The skews of the response curves towards the centroid, however, mean that F1 provides a sharper boundary than the centroid; indeed, for most of these vowels, the precise limits of the range of acceptable frequencies may not be at the centroid and F1 frequencies, but rather just inside the F1 boundary, and just outside the centroid boundary (or, slightly shifted towards higher frequencies for mid vowels, and towards lower frequencies for low vowels).

The sharp limit provided by F1 is not unexpected: F1 is a spectral peak, whereas there is no acoustic landmark corresponding to the centroid frequency. But this sharp limit seems to occur only with the peak associated with the nominal F1, and not with FN, no matter whether nominal F1 is above or below FN in frequency. In consequence, perceived nasal vowel height corresponds to a frequency that was higher than

the centroid for low nasal vowels, but lower than the centroid for all other vowels. In other words, the nominal F1 in nasal vowels (rather than the lowest-frequency spectral prominence) appears to contribute more to the perception of nasal vowel quality than would be predicted from its amplitude and bandwidth in a centroid measure.

II. EXPERIMENT 2: TWO-FORMANT AND ONE-FORMANT VOWELS

Experiment 1 shows that the perception of nasal vowel height is influenced both by FN and F1. These low-frequency peaks tend to differ in spectral prominence, and the more prominent peak appears to have the greater influence on the perceptual COG. One difference between oral and nasal vowels is that the low-frequency spectral peaks of certain oral vowels—such as F1 and F2 in back [ɑ] and [o]—may be equally prominent. We might speculate that two equally prominent peaks would contribute equally to the perceptual COG, as long as they are within 3.5 Bark of one another, and hence that listeners' responses to such oral vowels would be close to the centroid of the low-frequency region of those vowels. Indeed, the results of Chistovich et al. (1979) strongly support this speculation. Experiment 2 uses a design similar to that of Chistovich to further investigate the influence of spectral prominence on the perceptual COG.

A. Method

1. Stimuli

One- and two-formant stimuli that sounded like the vowels [ɑ], [o], and [u] were made on the parallel synthesizer at Haskins Laboratories. Each 360-ms stimulus had steady-state formant frequencies, with formant amplitude decreasing over the final 60 ms. The fundamental frequency of each stimulus increased from 110 Hz to 120 Hz over the initial 60 ms and then gradually decreased over the remainder of the vowel to 90 Hz.

In order to minimize the possibility of confounding a response bias towards a particular formant with the influences of formant amplitude, we attempted to synthesize each two-formant vowel approximation so that the two formants had equal amplitudes. This was possible for [ɑ] and [o], but not for [u], since attempts to equate peak amplitudes produced an unacceptable decrement in vowel quality. F1 and F2 bandwidths were set at 80 Hz for all three vowels.

Table 3 gives the parameter values for F1 and F2 frequencies of the two-formant vowels.

Table 3. *Two-formant vowels. Numbers represent the frequency (in Hz) of the two poles (F1 and F2) of the two-formant vowels, [ɑ], [o], and [u].*

	F1	F2
[ɑ]	780	1100
[o]	430	775
[u]	335	785

The difference between the frequencies of F1 and F2 was less than 3.5 Bark for all three vowels: 1.7 Bark for [ɑ], 2.5 Bark for [o], and 3.2 Bark for [u]. (The F2-F1 difference in naturally spoken American English /o/ and /u/ usually exceeds 3.5 Bark (Peterson and Barney, 1952; Syrdal and Gopal, 1986). While F1 and F2 frequencies for [o] of still fall within the normal range for American English, the F2 frequency of [u] had to be unusually low in order to have less than 3.5 Bark between F1 and F2. Consequently, this vowel sounded more back than is usual for American English, although subjects readily identified it as an exemplar of /u/.) DFT spectra of the two-formant vowels are given in Figure 6. This figure also shows the centroid of each vowel, which had a cutoff 5 dB below the trough between the two spectral peaks. This method of determining the frequency range of the centroid measure differed from the method used in Experiment 1, where the lower-amplitude cutoff was the lowest-amplitude point within a preselected frequency range. This frequency range was not appropriate for the two-formant vowels of this experiment. The cutoff of 5 dB below the trough between the two formants was chosen because it was likely to include that part of the spectrum that was most important auditorily, as discussed above for Experiment 1.

For each two-formant vowel, a series of one-formant vowels was synthesized. The center frequencies of the one-formant stimuli in each series varied in equal acoustic steps such that the range included F1 and centroid matches relative to the corresponding two-formant vowel, as seen in Figure 7. (See Table 4 for parameter values of these one-formant stimuli.) The formant-frequency increment between stimuli was 27 Hz for the *ɑ* series, 16 Hz for *o*, and 12 Hz for *u*. The bandwidth of all one-formant stimuli was set at 80 Hz.

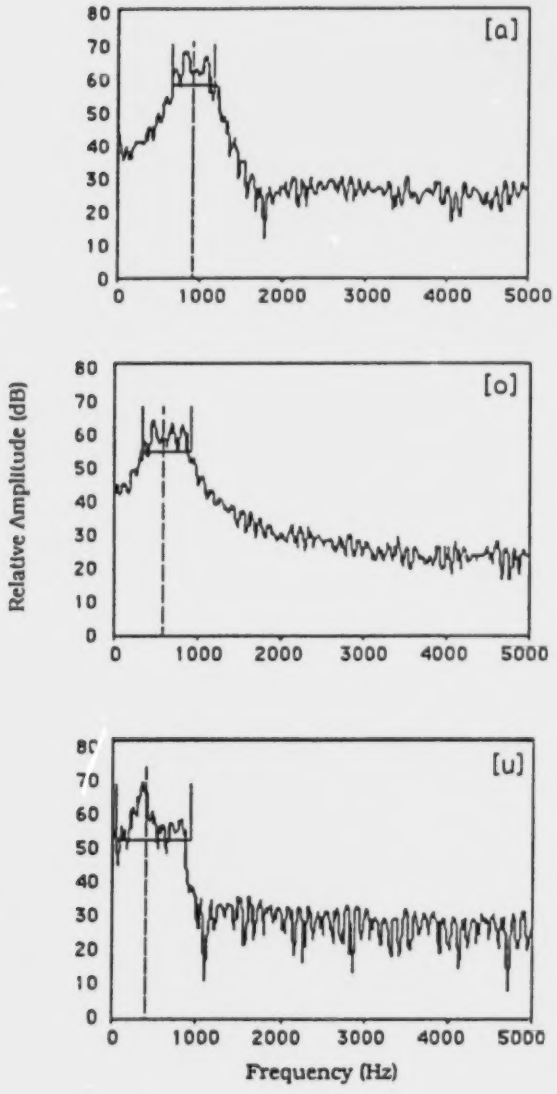


Figure 6. DFT spectra of the two-formant vowels of Experiment 2. The centroid measure, taken 5 dB below the trough between the two formants, is shown by the dashed line.

2. Subjects

20 students, aged 18 to 25 years, were paid for their participation in the experiment. 14 of these had served in the first experiment; as there was no difference between this group and the other 6 subjects, their responses were pooled.

3. Procedure

Each two-formant vowel was paired with each one-formant vowel from the appropriate series, with the two-formant reference vowel always being the second member of the pair. This resulted in three sets, *a*, *o*, and *u*, of one-formant—two-formant pairings. To avoid clustering of responses in the center of each set, two truncated versions were constructed from each of the three complete sequences: one omitted the first two members of the one-formant series and the other omitted the last two members. Because pilot tests showed that listeners found even the truncated sequences were too long, each truncated sequence was further shortened by removing the even-numbered members. The ordered sequences therefore consisted only of the odd-numbered stimulus pairs from each truncated set. (For all three vowel sets, these ordered sequences included the F1- and centroid-matched pairs.) Even-numbered stimulus pairs, although not present in the ordered sequences, occurred in "subsequences" which consisted of three adjacent odd-even-odd stimulus pairs (e.g., 5-6-7). All stimulus pairs (even- and odd-numbered) could also be heard as individual pairs.

In each test trial, listeners first heard an ordered sequence. They were told to determine that region of the sequence in which the vowel pairs sounded the most similar, and to listen to subsequences in that region. They were told that each subsequence contained an "intermediate" vowel pair (the even numbered pair). In addition to listening to sequences and the three-pair subsequences, subjects were advised to listen to individual pairs in the preferred region before making their decision. Other aspects of the procedure were as described for Experiment 1. As before, the listeners' task was to select that vowel pair from each set whose members were most similar in vowel quality. Two randomly ordered presentations of each of the two truncated versions of the three vowel sets gave a total of 12 trials per listener.

B. Results

Listeners varied considerably in the number of times the "intermediate" (i.e., even-numbered) vowel pairs were requested (either as part of a subsequence or as an individual pair) and, consequently, in the frequency with which these pairs were selected as best matches. For this reason, responses of the 20 listeners to each of the vowels *a*, *o*, and *u* were pooled across the four presentations of that set and then smoothed based on running means using a window of three stimuli.

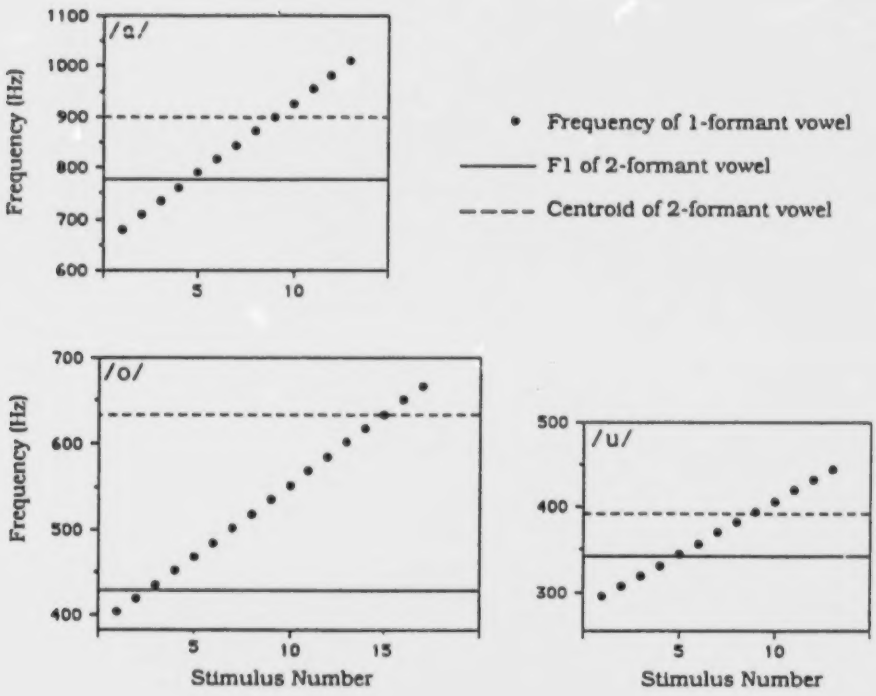


Figure 7. **F1 frequencies (solid dots) for the stimuli comprising the one-formant series of Experiment 2. F1 (solid line) and centroid (dashed line) frequencies of the corresponding two-formant vowel are shown for reference.**

Table 4. *One-formant vowels. Numbers represent the frequency (in Hz) of the single pole for the members of the three one-formant series.*

	Stimulus Number																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
a	680	707	735	762	790	817	845	872	900	927	955	982	1010	-	-	-	-
o	403	419	436	452	469	485	502	518	535	551	568	584	601	617	634	650	667
u	295	307	320	332	345	357	370	382	395	407	420	432	445	-	-	-	-

Figure 8 shows the smoothed responses to each vowel set. As in Experiment 1, the one-formant "best matches" chosen by listeners tended to fall between F1 and the centroid of the corresponding two-formant vowel. The mean stimulus number of these best-match responses (for each subject) differed significantly from the stimulus number of the F1-matched pair (for *a*, *o*, and *u*, $t(19) = 4.84$, 7.73, and 2.98 respectively, $p < .01$). Listeners' mean responses also differed significantly from vowel pairs matched for centroid frequency ($t(19) = 8.99$ (*a*), 12.95 (*o*), and 12.16 (*u*), $p < .01$). That is, the perceived quality of the two-formant vowels corresponded neither to their first formant frequencies nor to their centroid frequencies. However, subjects' matches were significantly closer to F1 frequency than to the centroid frequency for the non-low vowels *o* ($t(19) = 2.61$, $p < .05$) and *u* ($t(19) = 4.59$, $p < .01$).

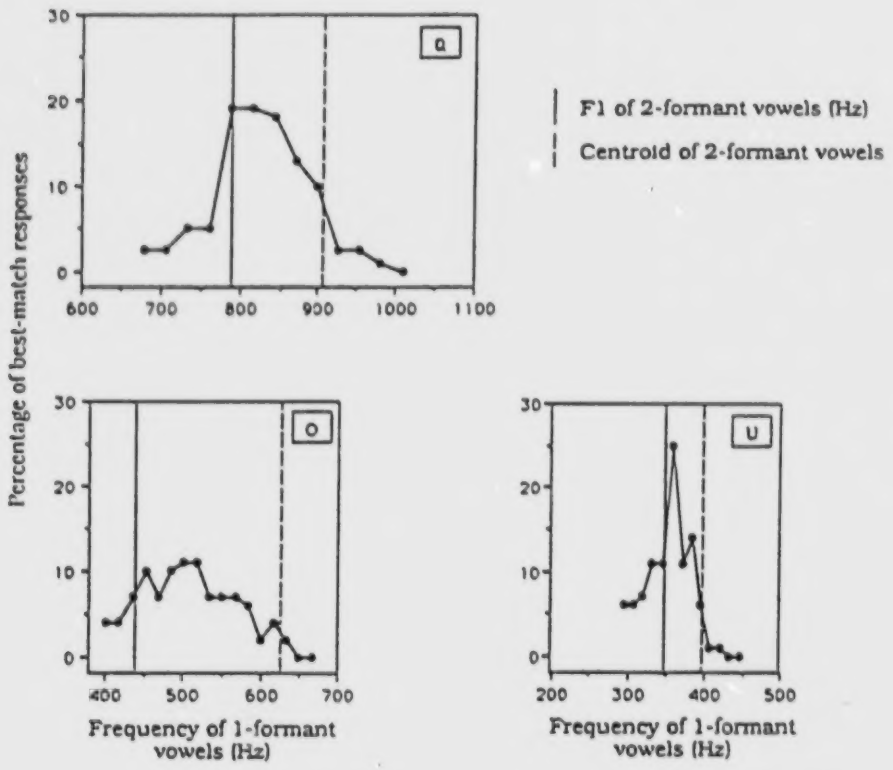


Figure 8. **Smoothed best-match responses to the pairings of one- and two-formant vowels for each vowel set of Experiment 2. F1 and centroid frequencies of the two-formant vowels are shown as vertical lines.**

C. Discussion

The center frequency of the one-formant stimuli which listeners judged as most similar in quality to a two-formant standard fell between the frequencies of the two formants of the standard. To this extent, these results for back oral vowels, with their two formants within 3.5 Bark of one another, are like those of Chistovich and like our data for nasal vowels from Experiment 1 in that they support the COG hypothesis: perceived vowel quality was not determined by the frequency of a spectral peak, but apparently depended on the combined influence of adjacent peaks in the vowel spectrum. But the results of Experiment 2 also differ from those of Experiment 1, as well as those of Chistovich.

First, although the center frequency of the one-formant vowels selected by listeners was significantly different from the F1 frequency of the two-formant vowels, these responses were also significantly closer to F1 frequency than they were to the centroid frequency for two of the three vowels, *o* and *u*. In the case of *ɑ*, this result cannot readily be attributed to a difference in relative prominence of the two spectral peaks. These data appear to differ from those of Chistovich et al. (1979), which showed that, when F1 and F2 amplitudes of two-formant vowels were identical, the frequency of the one-formant best-matches was about halfway between F1 and F2 frequencies. Our data do not support Chistovich et al.'s conclusions on the effect of formant amplitude on perceptual COG, but are more supportive of Klatt's (1985) conclusion that the influence of formant frequency overrides that of formant amplitude. In our two-formant data, it is the frequency of F1 that appears to

be the primary determinant of the perceptual COG.

The perceptual saliency of F1 in this experiment also contrasts with our findings for Experiment 1. Listener responses to [ɛ] from Experiment 1 and [o] from Experiment 2 illustrate this difference. For both of these mid vowels, the separation between the first two spectral peaks was 2.5 Bark, and these spectral peaks fell at roughly the same frequencies for both vowels (400-450 Hz for the first peak and 700-750 Hz for the second peak). But best-match responses to [ɛ] were closer to the centroid, while responses to [o] were closer to F1. That is, the nasal vowel showed a stronger COG effect, or more spectral averaging, than the oral vowel with the same spacing of low-frequency formants. This result holds for all vowels of Experiments 1 and 2: for all of the multiformant nasal vowels (except [ĩ], in which F1-FN separation exceeded 3.5 Bark), but for none of the two-formant oral vowels, perceived quality was nearer to the weighted average of the first two spectral peaks than to the frequency of F1.

This difference between the two experiments clearly contradicts our earlier speculation that the perceptual COG would be closer to the centroid in vowels with two equally prominent spectral peaks (as in two-formant [ɑ] and [o]) than in vowels whose low-frequency peaks differ in spectral prominence (as in the vowels of Experiment 1). One possible cause of the difference may be the presence of high-frequency influences in the multiformant nasal vowels, compared with their absence in the two-formant oral vowels. Of more immediate interest here, though, is the difference between the stimuli of Experiments 1 and 2 in the shape of the spectrum at low frequencies. As discussed earlier, the low-frequency spectra of nasal vowels are characterized by broad peaks separated by only shallow troughs, whereas the spectra of non-nasal oral vowels in the same frequency range have narrow peaks with well-defined troughs (e.g., Figure 4); the two-formant stimuli of Experiment 2 were of this second type (Figure 6). These data, then, point to the possibility that a broad low-frequency spectral prominence may be responded to by the auditory system in a different way from the more sharply defined spectral prominences of oral vowels. Specifically, we speculate that broader, flatter spectral prominences result in a stronger COG effect than do sharply defined spectral peaks. Experiment 3 was designed to test this hypothesis.

III. EXPERIMENT 3: WIDE AND NARROW BANDWIDTH VOWELS

A. Method

1. Stimuli

All stimuli in this experiment were two-formant vowel approximations in which degree of spectral prominence was varied by manipulating formant bandwidths. The design of the experiment was to pair a series of vowels having medium-bandwidth formants with two types of reference vowel, one having narrow-bandwidth formants and the other having wide-bandwidth formants.

The reference vowels were [o] and [ɑ], and each was synthesized with narrow- and wide-bandwidth versions: the bandwidth of F1 and F2 was 45 Hz in the narrow bandwidth versions and 150 Hz in the wide bandwidth versions. Table 5 gives the fundamental and formant frequencies of these reference vowels. To minimize the influence of interactions between harmonic and formant frequencies on the effective bandwidths within a stimulus, fundamental frequency remained constant throughout each vowel, with a value chosen to keep the difference between harmonic and formant frequencies as uniform as possible within all members of the medium-bandwidth series described below.

Table 5. *Narrow and wide bandwidth reference vowels. Numbers represent the frequency (in Hz) of the fundamental (F0) and of the two poles (F1 and F2) of the reference vowels.*

	F0	F1	F2
[o]	110	425	755
[ɑ]	115	700	1045

Medium-bandwidth versions of [o] and [ɑ], with F1 and F2 bandwidths of 75 Hz, were also generated. Each of these two medium-bandwidth vowels formed the midpoint of a series of medium-bandwidth vowels which was created by increasing or decreasing F1 and F2 frequencies in equal acoustic steps (25 Hz for *o* and 30 Hz for *ɑ*), keeping constant the absolute frequency difference between the two formants. Figure 9 shows the formant frequencies for the members of these medium-bandwidth series. Stimulus 5 in the *o* series and stimulus 4 in the *ɑ* series are the medium-bandwidth versions of the reference vowels and are therefore identical in centroid value to the references. Parameter values for all medium-bandwidth stimuli are given in Tables 6 and 7.

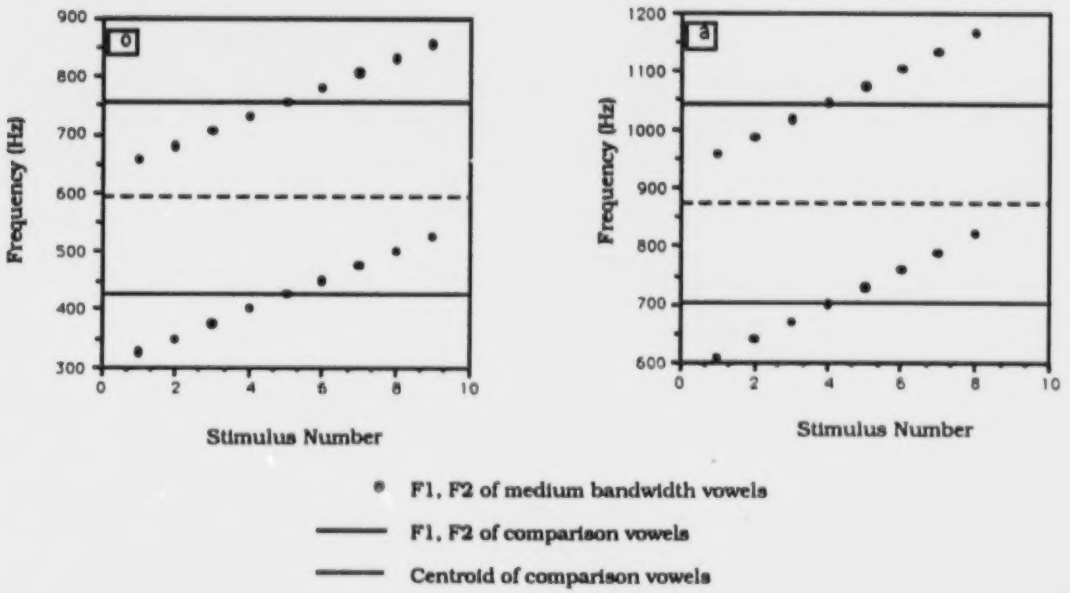


Figure 9. **F1 and F2 frequencies (solid dots) comprising the medium-bandwidth series for each vowel set of Experiment 3. F1 and F2 (solid lines) and centroid (dashed lines) frequencies of the corresponding narrow- and wide-bandwidth vowels are shown as horizontal lines.**

Table 6. *Medium-bandwidth o series. Numbers represent the frequency (in Hz) of the two poles, F1 and F2.*

	Stimulus number								
	1	2	3	4	5	6	7	8	9
F1	325	350	375	400	425	450	475	500	525
F2	655	680	705	730	755	780	805	830	855

Table 7. *Medium-bandwidth a series. Numbers represent the frequency (in Hz) of the two poles, F1 and F2.*

	Stimulus number							
	1	2	3	4	5	6	7	8
F1	610	640	670	700	730	760	790	820
F2	955	985	1015	1045	1075	1105	1135	1165

The stimuli as described so far were generated by parallel formant synthesis. In a second stage of synthesis, stimuli whose measured centroids fell more than 5 Hz from the mean frequency of F1 and F2 were modified by changing the amplitude of individual harmonics, using a harmonic synthesizer. Modifications were minor, and made so as to increase the symmetry of spectra as well as conformity to the desired centroid values. To control for synthesizer-specific effects, all stimuli were resynthesized using the harmonic synthesizer, although only some stimuli underwent harmonic amplitude modifications.

Figure 10 shows the two ways in which the centroid was measured. First, using the standard power spectrum with a decibel scale, the cutoff was set at 15 dB below peak formant amplitude, this being a point which captured all peak and trough information for all three bandwidth conditions, while avoiding the lower-amplitude skirts which were unlikely to affect perception. The second method used a linear power spectrum in order to find psychoacoustically valid and procedurally reliable frequency boundaries within which to calculate the centroid. It allowed us to check that the frequency boundaries delineated by the first method were appropriate. In the linear spectra, the power cutoff was at 0 and the frequency cutoffs were at the harmonic peaks with no more than four linear units' amplitude that were closest to each formant. The two types of centroid calculations gave values within 6 Hz of each other for each type of stimulus.

The measured step size between centroid frequencies (and also formant frequencies) of medium-bandwidth stimuli was 25 Hz for the *o* series and 30 Hz for the *a* series, plus or minus 5 Hz. Measured formant peaks in each stimulus differed by less than 2 dB. Each vowel was 360 ms long, with amplitudes adjusted so that all the stimuli sounded about equally loud in informal tests.

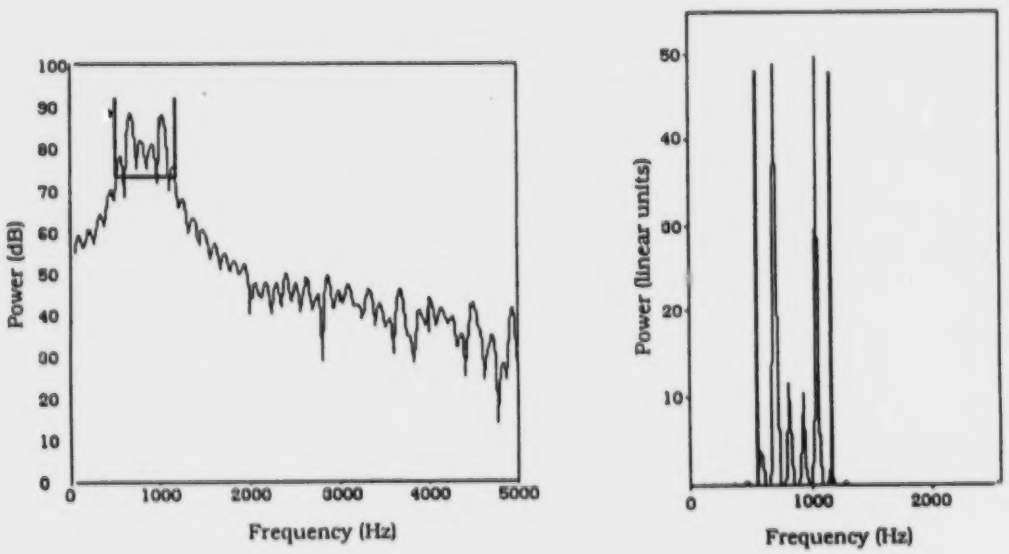


Figure 10. **Centroid measures used in Experiment 3. Left panel: centroid measured using a logarithmic power spectrum, with power cut-off at 15 dB below peak formant amplitude. Right panel: centroid measured using a linear power spectrum with power cut-off at zero.**

Figure 11 shows DFT spectra of the [o] and [*a*] reference vowels and the corresponding medium-bandwidth stimulus. The medium-bandwidth stimuli sounded nonnasal, but nevertheless slightly more nasal than the narrow-bandwidth stimuli. The wide-bandwidth stimuli sounded quite nasal. The manipulations of F1 and F2 affected the perceived height of vowels in the *o* series; in the *a* series, the main effect was in height for stimuli 1-3, and backness for stimuli 4-8.

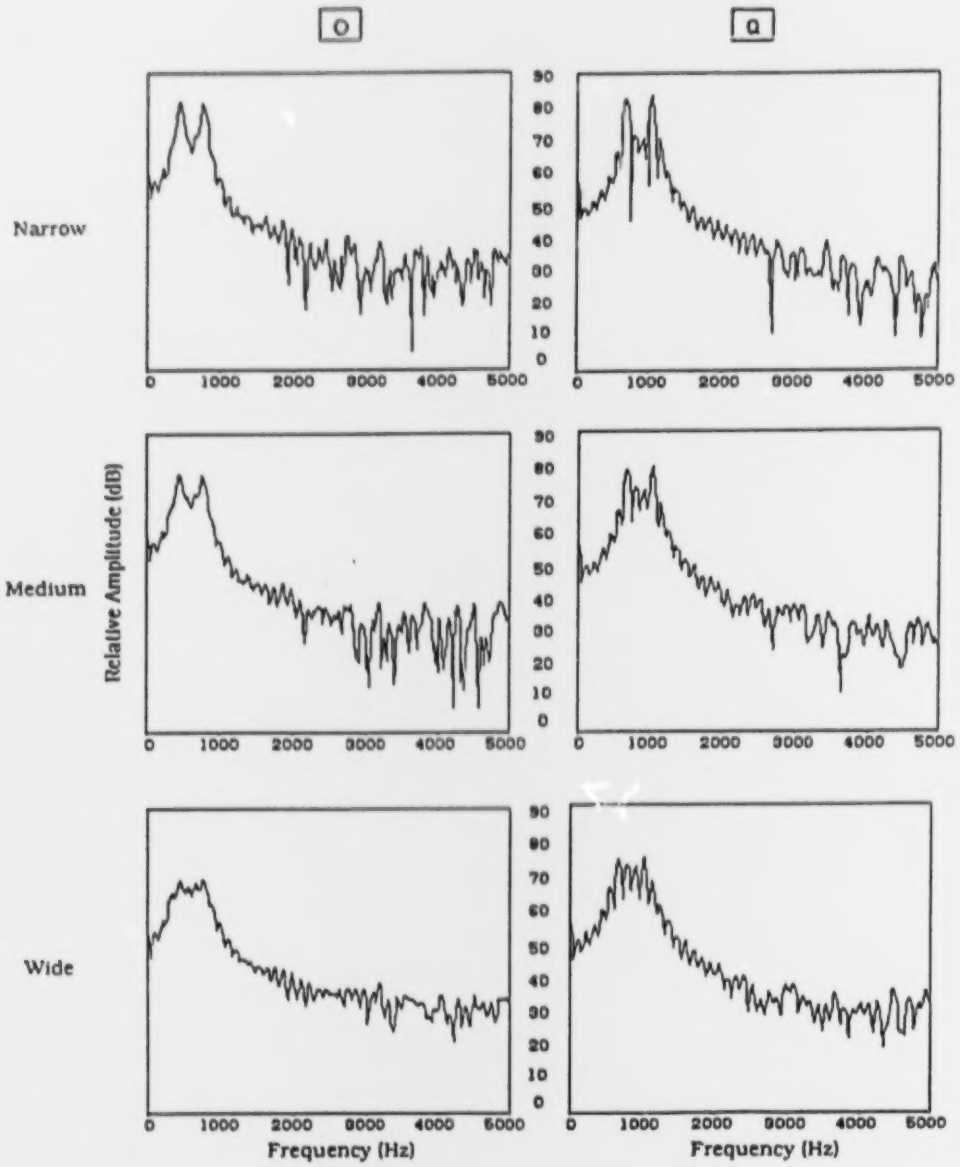


Figure 11. DFT spectra of the narrow-bandwidth (top panels) and wide-bandwidth (bottom panels) reference vowels of Experiment 3. The medium-bandwidth vowels (middle panels) have the same formant and centroid frequencies as the corresponding reference vowels.

2. Subjects and Procedure

The procedure was similar to that of Experiment 1 in that all stimuli occurred in the ordered sequences and there were no subsequences. The test sequences consisted of each narrow- or wide-bandwidth vowel, paired with each medium-bandwidth vowel from the appropriate series, giving four ordered sequences of vowel pairs, medium-narrow for *o* and *a*, and medium-wide for *o* and *a*. The medium-bandwidth vowel was always the first member of each pair. Listeners were tested on complete and truncated versions of these sequences, in which no, one, or two members of the medium-bandwidth series (taken from the beginning or end of each series) were omitted. Each of the resultant four vowel sets occurred four times in the test for a total of 16 sets presented in random order to each of 10 paid student volunteers. Six of the 10 subjects had participated in Experiments 1 and 2.

B. Results

The responses of the 10 listeners are shown in Figure 12. Responses were not affected by whether listeners had served in the previous two experiments, so the data are pooled over subjects. The patterns of responses to the narrow-bandwidth versions of [o] and [ɑ] (upper panels) were similar in that the perceived quality of the narrow-bandwidth stimulus corresponded to the medium-bandwidth stimulus with the same centroid frequency (indicated by the dashed line). In the case of the *a* series, the majority of choices were of the centroid-matched pair and the remaining ones were nearly evenly distributed between the adjacent two stimulus pairs.

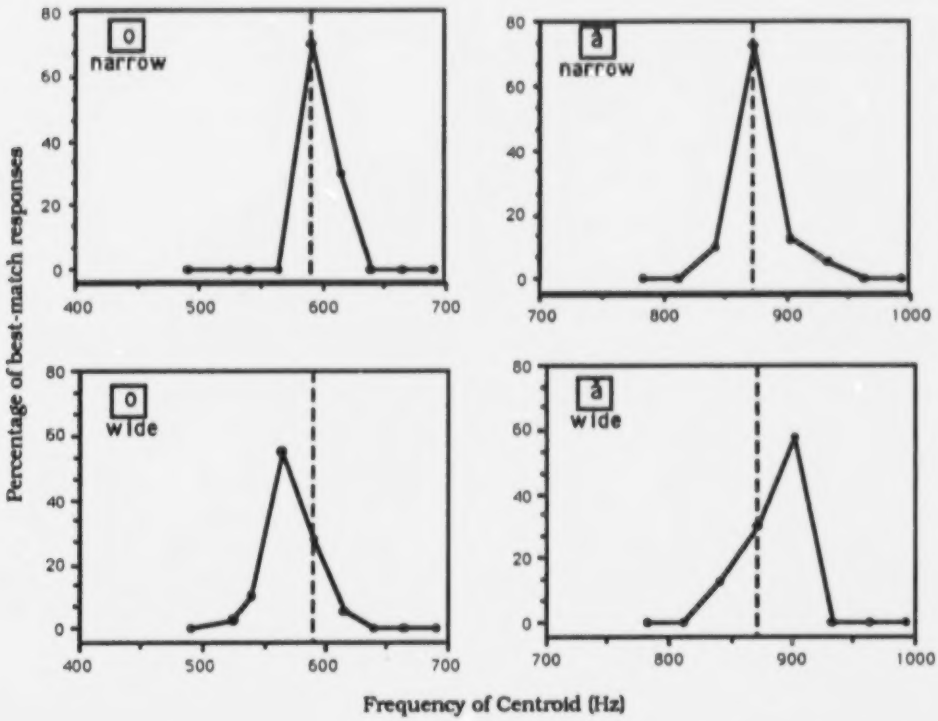


Figure 12. **Percent best-match responses to the medium-bandwidth vowels paired with the narrow-bandwidth references (top row) and wide-bandwidth references (bottom row) of Experiment 3. The dashed line represents the centroid frequency of the reference vowel of each set.**

For *o*, the centroid-matched pair also accounted for most of the responses. But the remaining responses to this vowel had a higher centroid frequency than that of the reference, indicating a tendency for the narrow-bandwidth [o] to sound lower relative to the medium- bandwidth vowels than its formant frequencies lead us to expect. Due to this skew towards higher frequencies, the difference between listeners' responses and the centroid-match stimulus number for narrow-bandwidth [o] was significant ($t(9) = 2.88$, $p < .05$). This difference was not significant, as Figure 12 suggests, for narrow-bandwidth [ɑ] ($t(9) = 0.83$).

The responses to wide-bandwidth versions of [o] and [ɑ] (lower panels in Figure 12) had a very different pattern. For these vowel sets, wide- and medium-bandwidth vowels with the same formant and centroid frequencies were not chosen as the most closely matched in vowel quality. For the *a* series, the centroid of the best-match medium-bandwidth stimulus was 30 Hz higher than the centroid of the wide-bandwidth reference. For *o*, the shift was in the opposite direction. Both shifts were significant at $p < .$ ($t(9) = 3.86$ for *a* and 5.67 for *o*).

C. Discussion

The hypothesis that a broad, flat spectral prominence leads to a higher perceptual COG was not confirmed, but such spectra did seem to be heard in a qualitatively different way from spectra with sharper formant peaks. The medium-bandwidth stimuli judged as most similar in quality to the narrow-bandwidth comparisons were those that shared the same peak (and centroid) frequencies, whereas in comparisons involving wide-bandwidth stimuli, the preferred stimulus was one step below (for *o*) or above (for *a*) the centroid match.

This bidirectional shift from the centroid-matched stimulus appears to result from listeners evaluating perceived quality in terms of similarity of spectral shape (as well as approximate correspondence in frequency of spectral peaks). Shifts in opposite directions occurred as a result of an experimental artifact, but nevertheless support our hypothesis that the wide and narrow bandwidths are responded to in different ways by the perceptual system. Despite stimuli being constructed so that the variation in distance between harmonic and formant frequency was as small as possible across the medium-bandwidth stimulus sequence, some variation was unavoidable. For most stimuli, only one harmonic was principally excited for each formant, but for some, two harmonics per formant were excited; there were, of course, intermediate degrees of variation in other stimuli. Just this type of variation occurred among the stimuli near the centroid value of the reference stimuli.

Calculations from the formant and fundamental frequencies given in Tables 5, 6, and 7 show that the variation in the medium-bandwidth series between the centroid match and its two adjacent stimuli was as follows. For *o*, the fourth and seventh harmonics fall 40 Hz, 15 Hz, and 10 Hz from F1 and F2 frequencies for stimuli 4, 5, and 6 respectively. For *a*, the sixth and ninth harmonics fall 20, 10, and 40 Hz from F1 and F2 for stimuli 3, 4, and 5 respectively. The effects of this variation can be seen in Figure 13. Figure 13a shows, for the *o* series, the narrow- and wide-bandwidth reference stimuli, and the three stimuli from the medium-bandwidth sequence that most closely corresponded in centroid frequency with the references (the centroid match stimulus 5, and stimuli 4 and 6). The corresponding data for *a* are shown in Figure 13b. The three medium-bandwidth spectra in each panel differ in the breadth of their effective bandwidths, and those stimuli with the largest difference between harmonic and formant frequency have the broadest and flattest spectral envelopes—stimulus 4 for *o*, and stimulus 5 for *a* These are precisely the stimuli that were preferred in the listening test: the difference between stimuli in spectral envelope explains the bidirectionality of the response data.

Evidently, then, listeners sacrificed some precision in terms of peak or centroid frequency in order to maximize similarity in overall spectral shape, which corresponded to perceived similarity in vowel quality. Informal tests lead us to believe that there will be strict limits on the frequency differences that will be tolerated in order to attain a better match of spectral shape, but that, within these limits, overall spectral shape is more important than peak frequency in determining similarity in vowel quality. We make no claims about the qualities to which listeners were responding. Our manipulations of spectral shape tended to affect perceived nasality, and there were also other effects, such as "brightness" of timbre. Listeners may have matched for degree of nasality in the *a* series, since the preferred medium-bandwidth stimulus (stimulus 5) does sound more nasal than the centroid match (stimulus 4). But the preferred and centroid-match stimuli in the *o* series (stimuli 4 and 5) sound about equally nasal, which indicates that listeners responded to some other quality in this series.

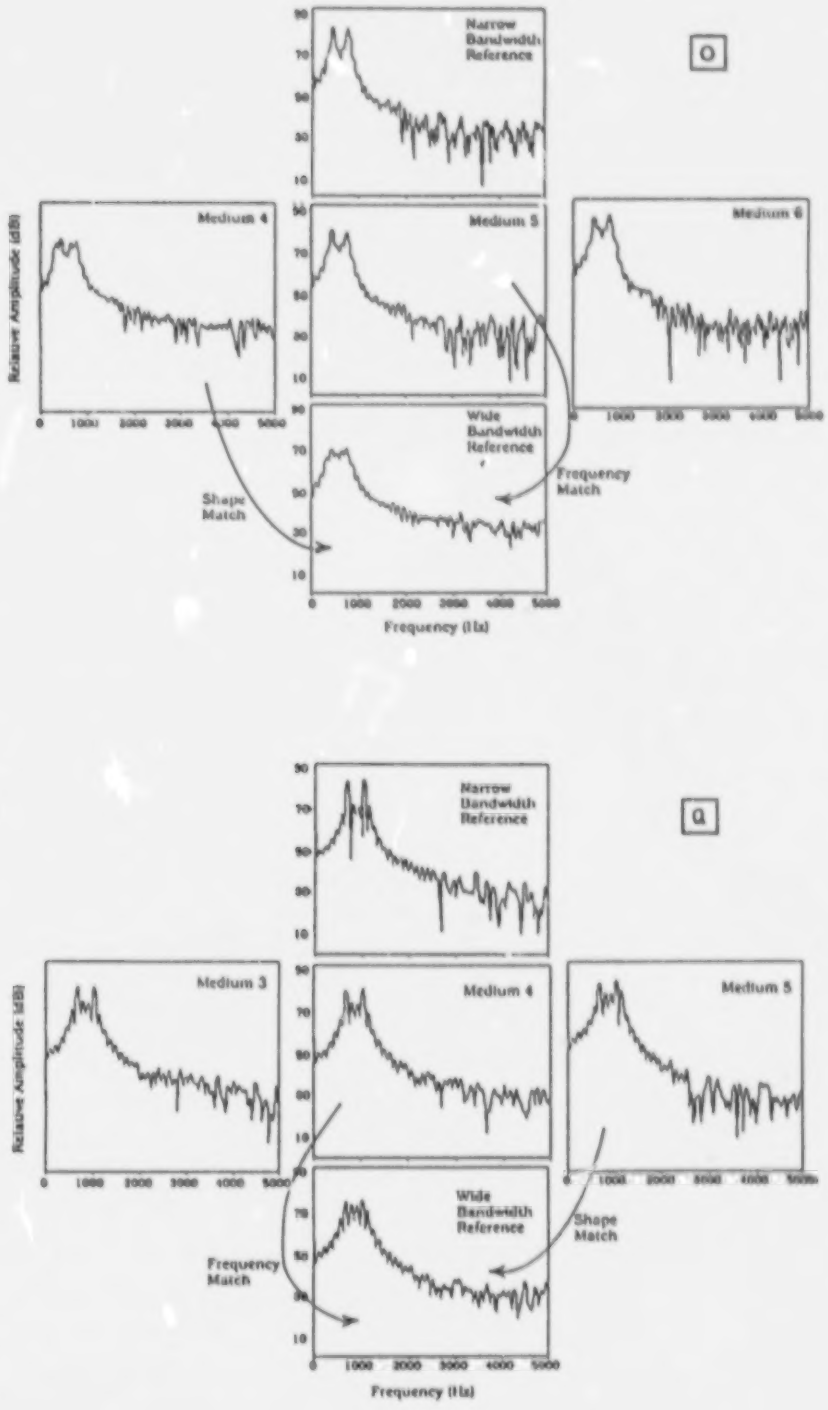


Figure 13. **DFT spectra of the medium-bandwidth vowels that were closest to the narrow- and wide-bandwidth references in centroid frequency in Experiment 3. For [o] (part a) stimulus 5 of the medium-bandwidth series was the centroid match, but stimulus 4 was the closest match in spectral shape. For [a] (part b) medium-bandwidth stimulus 4 was the centroid match, but stimulus 5 was the closest shape match.**

Our hypothesis that the spectral envelope is less important when formant peaks are well defined is supported by the finding that narrowing formant bandwidth did not affect the perceptual COG of [o] and [ɑ]. Nevertheless, it may be possible to explain the asymmetry of the responses to narrow-bandwidth [o] in terms of bandwidth differences. For this vowel, stimulus 6 was the only stimulus other than 5, the centroid match, to receive any votes for "best match," and both had only small differences between harmonic and formant frequency—15 Hz for stimulus 5, and 10 Hz for stimulus 6. These two stimuli contrast markedly with stimulus 4, which has a 40 Hz harmonic-formant difference and consequent excitation of two harmonics per formant (see Figure 13). On the other hand, the response curve for [ɑ] shows no such asymmetry: there was the same negligible number of responses to stimulus 3, which had a 20 Hz harmonic-formant difference, as to stimulus 5, with a 40 Hz difference. If the shape of the spectral envelope is important in judging quality when peaks are sharply defined, then the boundary between what is comparable and what is not appears to operate within a very narrow range.

IV. GENERAL DISCUSSION

As outlined in the Introduction, previous studies indicate that though a center-of-gravity effect influences the perception of vowel quality, formant frequency contributes more heavily to that perceptual COG than "spectral averaging" would suggest. The three experiments reported here manipulated the prominence of the low-frequency region of vowel spectra and investigated the effect of these manipulations on the relative contribution of peak frequency and overall spectral envelope to perceived vowel quality. Their results suggest that the influence of the spectral envelope on perceived quality (especially perceived height) increases as the prominence of low-frequency spectral peaks decreases. Thus the perceived quality of vowels with well-defined peaks, such as the oral vowels of Experiment 2, was dominated by the frequency of F1. In contrast, perception of the non-high nasal vowels of Experiment 1, which had relatively broad peaks and shallow troughs in the F1 region, corresponded more closely to the spectral COG than to the frequency of a spectral peak. Similarly, in Experiment 3, the perceived quality of the stimuli with narrow formant bandwidths was determined by formant frequencies, while that of wide-bandwidth peaks depended on a combination of formant frequency and bandwidth.

While the general pattern of listeners' responses to our stimuli suggests an inverse relation between the degree of spectral prominence and the COG effect, the more detailed characteristics of these responses have not been adequately described. In Experiment 3, we explained the finding that listeners' responses to wide bandwidth [o] and [ɑ] were in opposite directions (relative to the centroid) in terms of the effective formant bandwidths. Similar explanations for the particular stimuli chosen by listeners in Experiments 1 and 2, though, are not possible: for most vowel sets, the spectral characteristics of the vowels belonging to the best-match pairs were not more similar than those of vowels belonging to less-preferred pairs.

The lack of a consistent spectral description of the basis for our listeners' decisions may be due to the use of linear (Hz) frequency measures and decibel units of amplitude. The auditory system imposes a number of transformations on the incoming signal which can enhance particular similarities and differences between speech signals (e.g., Delgutte, 1984; Seneff, 1985). Although many of these transformations are controversial, and much remains to be discovered, some are widely accepted. Among amplitude transformations, phon and sone scales are often used, but we retained the conventional decibel and linear units since the use of amplitude transformations (especially sones) is controversial. Frequency-domain transformations reflecting the critical bands of the peripheral auditory system are among the least controversial, however, and are particularly relevant to the perception of vowel quality. Indeed, Chistovich's formulation of the centroid uses the Bark frequency scale, which is essentially a critical-band measure.

Recalculation of our measures using the Bark scale left our results and conclusions unchanged since the Bark scale is almost linear at frequencies less than 500 Hz. An alternative transformation based on the equivalent rectangular bandwidth, or ERB-scale, of Moore and Glasberg (1983) gave more promising results. The ERB transformation differs from Bark and critical-band-rate metrics primarily in that it is somewhat (but not strongly) nonlinear at frequencies less than 500 Hz. Excitation patterns derived for complex tones using the ERB-rate scale resolve individual harmonics only at low frequencies, and produce some changes in overall spectral tilt. The ERB scale was attractive because its low-frequency nonlinearity was likely

to give the most opportunity for the appearance of significant differences from the standard linear-frequency spectra. Spectra were calculated for frequencies less than 2 kHz.

ERB-scale spectra from relevant pairs of stimuli in each of the three experiments were compared. Although these comparisons did not yield a simple picture, more consistencies emerged than were found in comparisons of other spectra. We identified two properties in which the reference stimulus was spectrally more similar to listeners' best matches than to less-preferred stimuli: first, the most prominent one or two harmonics were identical or closely similar in frequency and amplitude; and second, the overall shape of the spectrum in the immediate vicinity of these prominent harmonics was similar. Often, this similarity in overall shape was pronounced for either the left- or the right-hand skirt, but not for both. (It was usually not possible for both skirts to have similar slopes because, except in Experiment 3, pairs differed in the number of peaks in a given frequency range.)

Figures 14, 15, and 16 provide examples of ERB-rate spectra from Experiments 3, 2, and 1 respectively. Figure 14 gives ERB-rate spectra for the three most preferred medium-bandwidth vowels (thin lines) from the *o* set (top row) and *a* set (bottom row) of Experiment 3. The corresponding wide-band reference vowel (thick lines) is superimposed on each spectrum. Stimulus 4 in the *o* set and stimulus 5 in the *a* set are both more like the reference-vowel spectra than are the other two, less-preferred, stimuli. (The similarities and differences between the ERB-rate spectra are comparable to those discussed for the DFT spectra in Figure 13.)

An example of similar data for Experiment 2 is given in Figure 15. Stimulus 5 was the most, and stimulus 9 the least preferred of the three stimuli shown from the *a* set (see Figure 8). Stimulus 5 corresponds most closely with the two-formant reference both in most prominent harmonic and lefthand skirt.

For the four oral-nasal vowel pairs of Experiment 1 which had less than 3.5 Bark between F1 and FN, a similar description works well for *o*, *a*, and *æ*, but less well for *e*. ERB-rate spectra for *o* and *e* are shown in Figure 16. (Stimulus 5 was preferred most often for *o*, and stimulus 6 for *e*—see Figure 5.) The common spectral property of all four oral-nasal sets is that the most prominent peak of each preferred stimulus has two prominent harmonics rather than one. Similarity of skirts appears to be of secondary importance in that it does not compensate for a narrow major prominence (cf. stimulus 5 versus 6 for *e*), but it does increase the perceptual similarity where the major prominence is broad (cf. stimulus 5 versus 6 for *o*). This dual criterion is very similar to the one listeners apparently used in Experiment 3.

Our general conclusion, then, is that where harmonics are sufficiently prominent, their frequency is of prime importance to perceived vowel quality, and spectral shape is of secondary importance. When the lowest-frequency spectral prominence is broad, what is important is not so much the absolute frequency of some peak, nor yet of a spectral location derived by a spectral-averaging (COG) mechanism, but the general correspondence in shape of the entire region of spectral prominence (presumably within appropriate amplitude limits). No one peak can be said to dominate such a spectrum, and the auditory response is therefore also unlikely to be dominated by a narrow range of frequencies. This conclusion explains the apparent discrepancies in some of our data: the best-match stimulus chosen from a vowel series may have either a peak center frequency, or a centroid, above or below that of the reference stimulus, depending on the characteristics of the particular set of stimuli from which the choice is made.

A model of the perception of vowel quality must be able to deal with nasal vowels as well as non-nasal vowels. It follows that the model should include mechanisms that will distinguish between the internal representations of sounds with different degrees of spectral prominence by appropriately weighting their physical properties. That is, we should seek auditory transforms that will distinguish the poorly defined prominences and shallow troughs of nasal vowels from the sharp peaks and deep troughs characteristic of many oral vowels. One such transform could be enhancement of prominent spectral peaks attributed to lateral suppression (Houtgast, 1977; Moore & Glasberg, 1981), and consistent with the physiological data of Sachs and Young (1980) on synchrony, or phase-locking, of auditory nerve fiber responses to prominent spectral peaks. Sidwell and Summerfield (1985) have shown that effects indicative of suppression occur with vowel-like sounds, and are seen most clearly at frequencies less than 2 kHz—those most strongly connected with the perception of phonetic quality.

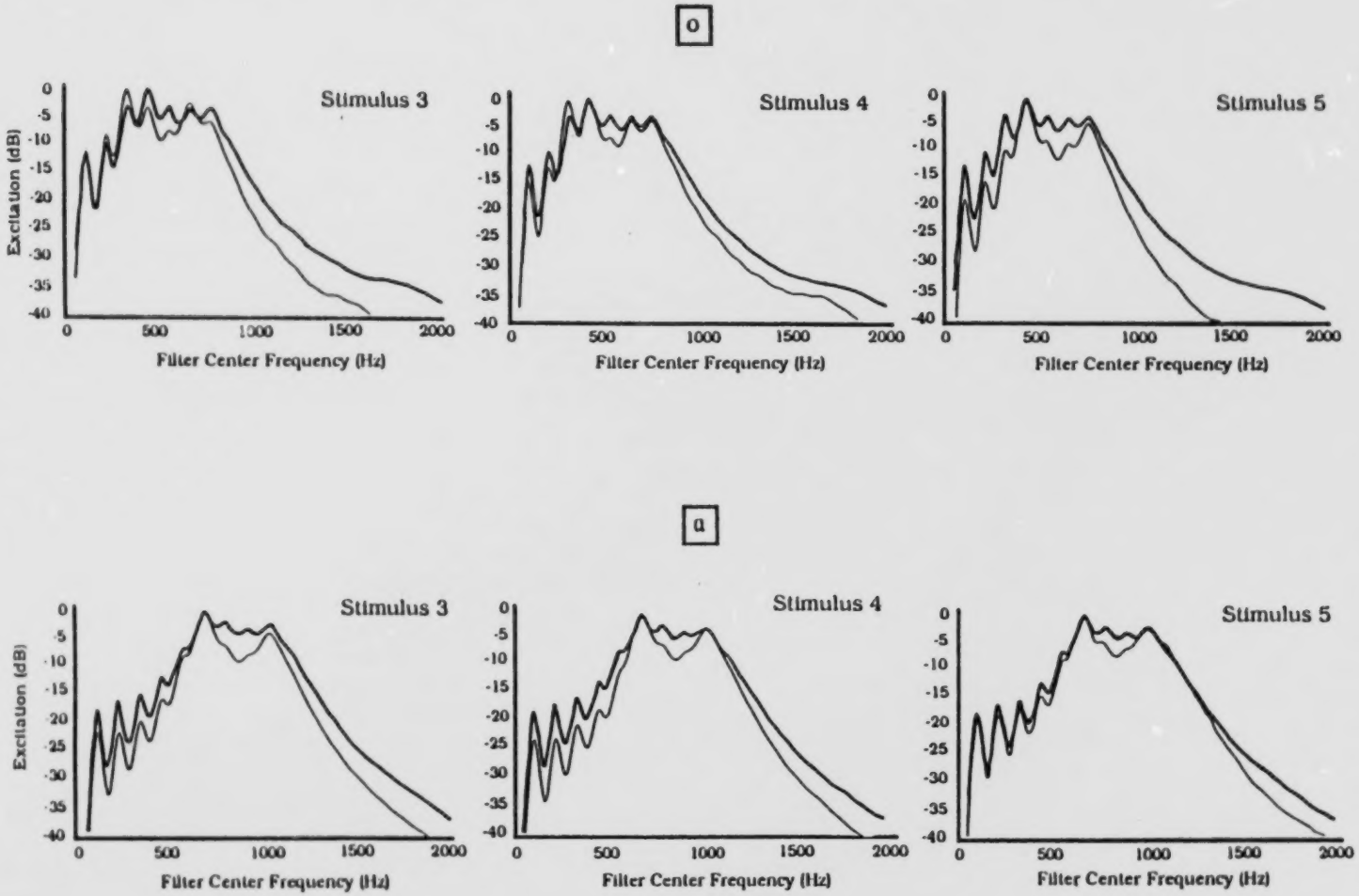


Figure 14. **ERB-rate spectra of the three most preferred medium-bandwidth vowels (thin lines) from each vowel set of Experiment 3. Each panel also shows the spectrum of the corresponding wide-bandwidth reference vowel (thick lines). As in the DFT spectra of Figure 13, medium-bandwidth stimulus 4 was the closest match in spectral shape for [o] (top panels), and stimulus 5 was the closest shape match for [A] (bottom panels).**

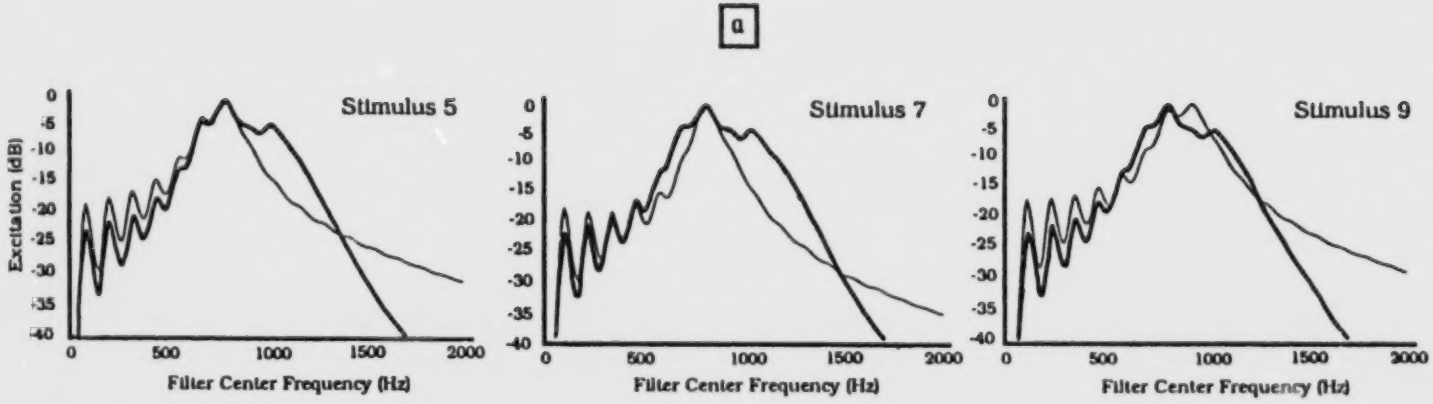


Figure 15. **ERB-rate spectra of three of the most preferred one-formant vowels (thin lines) from the A set of Experiment 2. Each panel also shows the spectrum of the two-formant reference vowel (thick lines). Stimulus 5 was the closest shape match in terms of both the most prominent harmonic and the lefthand skirt.**

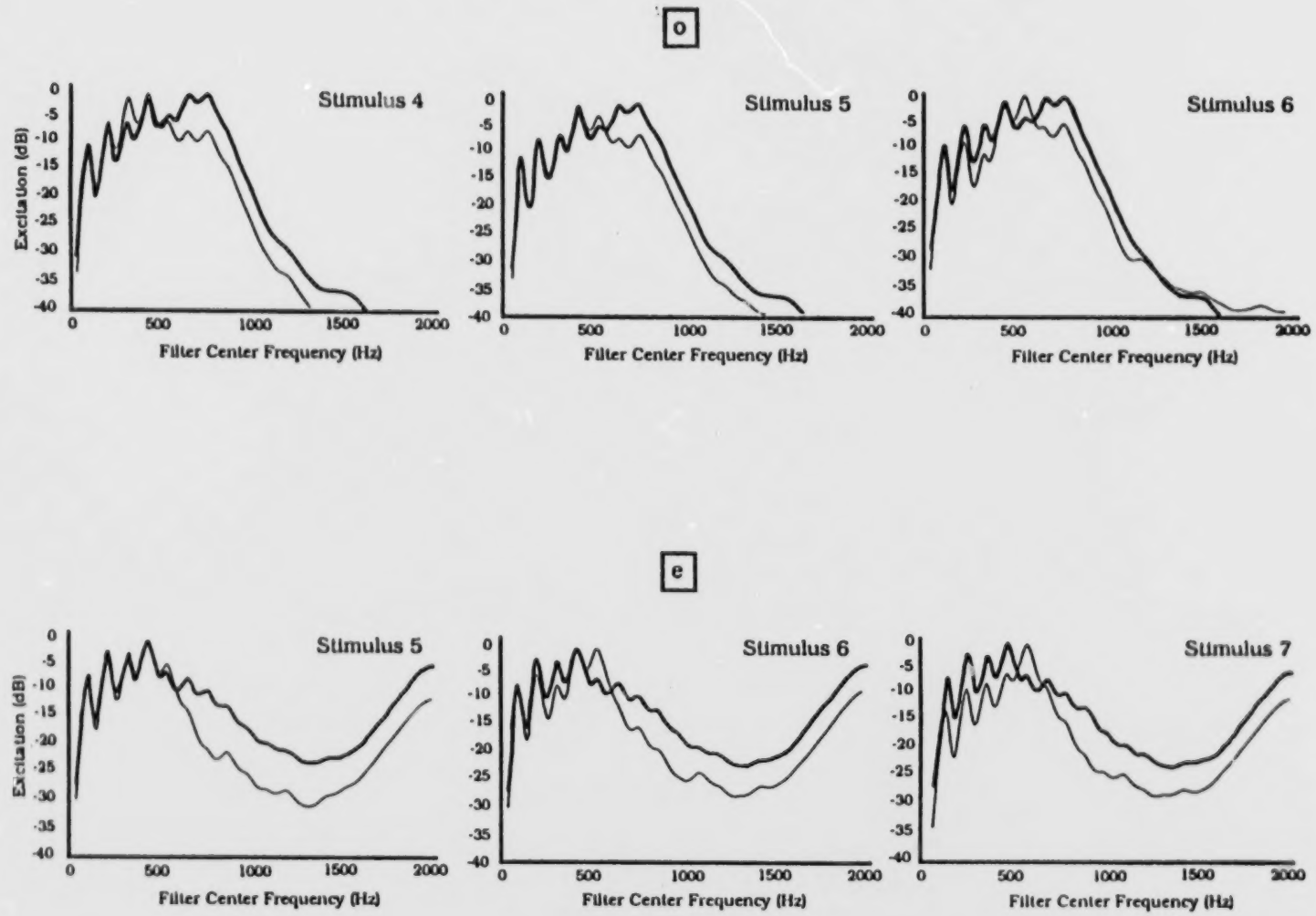


Figure 16. **ERB-rate spectra of the three most preferred oral vowels (thin lines) from the o and e vowel sets of Experiment 1. Each panel also shows the spectrum of the nasal vowel reference (thick lines). Stimulus 5 was the closest shape match for both vowel sets.**

The selective applicability of processes like phase-locking and suppression to physical signals differing in spectral prominence may underlie many of the apparent contradictions in the literature on the perception of vowel quality. For example, working largely with vowels whose physical spectra had well-defined peaks, Klatt (1985), like many others, emphasized the importance of center formant frequency to vowel quality. In earlier work, however, he found that listeners' judgments of the phonetic distance between synthetic vowel pairs strongly correlated with a metric of slope based on differences in slope near the peaks of critical-band spectra (Klatt, 1981; 1982). The spectral slope metric was sensitive to variation in parameters such as the number of critical band filters, their bandwidths, and aspects of the weighting function. All of these affect, or are reflected in, the amount of spectral contrast likely to appear in the internal representation of the sound.

Simple peak-peaking and center-of-gravity models of vowel perception fail not only for the poorly defined prominences of nasal vowels, but apparently also for back vowels. Assmann (1985) has suggested that details of the shape of the spectrum in the F1-F2 range and above make significant contributions to a back vowel's quality. This conclusion agrees exactly with ours, the spectra of back vowels representing an intermediate case between the clearly defined peaks of front vowels and the markedly broadened and attenuated low-frequency prominences of nasal vowels.

We also question the applicability of peak-picking models to the speech of women, children, and others who speak with a higher fundamental frequency than the average adult man, but who are probably not less intelligible (Verbrugge, Strange, Shankweiler, & Edman, 1976; but see Sundberg & Gauffin, 1982). Speech with a high F0 is likely to have more poorly-defined peaks, and the higher formant frequencies that often co-occur with high F0 are also likely to be less well-resolved by the auditory system.

We suggest, then, that peak-picking occurs only for certain spectra, perhaps due to auditory spectral enhancement taking place automatically through processes like phase-locking of auditory nerve fibers to a dominant frequency, and lateral suppression. Physical spectral without well-defined peaks will be responded to such that the properties of more extensive regions of the spectrum are given greater weight, because the lack of prominent spectral peaks will allow auditory nerve fibers to fire at their characteristic frequencies. Thus exclusively peak-picking models are appropriate only for certain groups of vowels—most clearly, front oral vowels, especially if spoken by adult males. But even for vowels with well-defined spectral peaks, aspects of spectral shape probably influence the perception of vowel height. Support for this view is provided by Darwin and Gardner's (1985) finding that phonemic identity is affected by changes in the amplitudes of harmonics which are relatively remote from the formant peak, and by Bladon and Lindblom's (1981) model of vowel quality based on whole-spectrum transforms. It may be that peak-picking models seem appropriate for front vowels simply because, for sounds with sharp spectral peaks that are well separated from other peaks, overall spectral shapes in the vicinity of the peak tend to be similar, at least for the first 10-15 dB below the peaks, as long as the frequencies of the most prominent harmonics are the same. Thus there is little opportunity to observe the influence of spectral shape for the first formants of front vowels.

The variability in spectral shape across all vowels and potential speakers raises the question of whether it makes sense to search for a single spectrally based measure of the perceptual center of gravity. The fact that we had to use different centroid measures in our experiments in order to stay close to Chistovich's formulation underscores the problem. Models of vowel perception that combine peak-picking with some sort of whole-spectrum approach are likely to be most successful, as Bladon and Lindblom (1981) and others have pointed out.

In summary, the data we have presented indicate that a comprehensive model of the perception of vowel quality (particularly height) will include transforms that produce different types of auditory response dependent upon the relative spectral prominence of the formant peaks. When a spectral peak is prominent, its frequency will dominate the response, whereas the overall spectral shape in the vicinity of a formant will become more influential in determining vowel quality as the formant itself becomes less well defined. Physiological processes that could underlie this type of response include phase-locking of hair-cell responses and lateral suppression. The hypothesis of spectral averaging for formants within 3.5 Bark may largely reflect the fact that formants that are close in frequency tend to be less spectrally (and hence auditorally) prominent (in that they are separated by relatively shallow troughs) than formants that are

widely spaced in frequency. The implication is that neither simple center-of-gravity nor peak-picking measures can be expected to account on their own for responses to more than a restricted set of vowels, because the perceptual quality of any given vowel depends in complex ways on its detailed spectral shape, including, of course, on the frequencies of any prominent formants.

REFERENCES

Ainsworth, W. A., & Millar, J. B. (1972). The effect of relative formant amplitude on the perceived identity of synthetic vowels. *Language and Speech, 15*, 328-341.

Assmann, P. F. (1985). *The role of harmonics and formants in the perception of vowel quality*. Unpublished doctoral dissertation, University of Alberta.

Assmann, P. F., & T. M. Nearey. (1987). Perception of front vowels: The role of harmonics in the first formant region. *Journal of the Acoustical Society of America, 81*, 520-534.

Beddor, P. S. (1982). *Phonological and phonetic effects of nasalization on vowel height*. Doctoral dissertation, University of Minnesota, Minneapolis, MN (Reproduced by Indiana University Linguistics Club, 1983).

Beddor, P. S., Krakow, R. A., & Goldstein, L. M. (1986). Perceptual constraints and phonological change: A study of nasal vowel height. *Phonology Yearbook, 3*, 197-217.

Bedrov, Ya. A., Chistovich, L. A., & Sheikin, R. L. (1978). Frequency position of the center of gravity of formants as a useful feature in vowel perception. *Soviet Physics Acoustics, 24*, 275-278.

Bhat, D. N. S. (1975). Two studies on nasalization. In C. A. Ferguson, L. M. Hyman, & J. J. Ohala (Eds.), Nasálfest: *Papers from a Symposium on Nasals and Nasalization* (pp. 27-48). Stanford, CA: Language Universals Project, Stanford University.

Bladon, R. A. W. (1983). Two-formant models of vowel perception: shortcomings and enhancements. *Speech Communications, 2*, 305-313.

Bladon, R. A. W., & Fant, G. (1978). A two-formant model and the cardinal vowels. *Kungl Tekniska Högskolan: Speech Transmission Laboratories-Quarterly Progress and Status Report, 1*, 1-8.

Bladon, R. A. W., & Lindblom, B. J. (1981). Modeling the judgment of vowel quality differences. *Journal of the Acoustical Society of America, 69*, 1414-1422.

Carlson, R., Fant, G., & Granström, B. (1975). Two-formant models, pitch, and vowel perception. In G. Fant & M. Tatham (Eds.), *Auditory analysis and perception of speech* (pp. 55-82). New York: Academic.

Carlson, R., & Granström, B. (1979). Model predictions of vowel dissimilarity. *Kungl Tekniska Högskolan: Speech Transmission Laboratories-Quarterly Progress and Status Report, 3/4*, 84-104.

Carlson, R., Granström, B., & Fant, G. (1970). Some studies concerning perception of isolated vowels. *Kungl Tekniska Högskolan: Speech Transmission Laboratories-Quarterly Progress and Status Report, 2/3*, 19-35.

Carlson, R., Granström, B., & Klatt, D. (1979). Vowel perception: The relative perceptual salience of selected acoustic manipulations. *Kungl Tekniska Högskolan: Speech Transmission Laboratories-Quarterly Progress and Status Report, 3/4*, 73-83.

Chistovich, L. A. (1985). Central auditory processing of peripheral vowel spectra. *Journal of the Acoustical Society of America, 77*, 789-805.

Chistovich, L. A., & Lublinskaya, V. V. (1979). The 'center of gravity' effect in vowel spectra and critical distance between the formants: Psychoacoustical study of the perception of vowel-like stimuli. *Hearing Research, 1*, 185-195.

Chistovich, L. A., Sheikin, R. L., & Lublinskaya, V. V. (1979). 'Centres of gravity' and spectral peaks as the determinants of vowel quality. In B. Lindblom & S. Öhman (Eds.), *Frontiers of speech communication research* (pp. 143-157). New York: Academic.

Darwin, C. J., & Gardner, R. B. (1985). Which harmonics contribute to the estimation of formant frequency? *Speech Communication, 4*, 231-235.

Delattre, P. (1954). Les attributs acoustiques de la nasalité vocalique et consonantique. *Studia Linguistica, 8*, 103-108.

Delattre, P., Liberman, A., Cooper, F., & Gerstman, L. (1952). An experimental study of the acoustic determinants of vowel colour; Observations on one- and two-formant vowels synthesized from spectrographic patterns. *Word, 8*, 195-210.

Delgutte, B. (1984). Speech coding in the auditory nerve II: Processing schemes for vowel-like sounds. *Journal of the Acoustical Society of America, 75*, 879-886.

Fant, G. (1960). *Acoustic theory of speech production*. The Hague: Mouton.

Fujimura, O., & Lindqvist, J. (1971). Sweep-tone measurements of vocal-tract characteristics. *Journal of the Acoustical Society of America, 49*, 541-558.

Fujisaki, H., & Kawashima, T. (1968). The roles of pitch and higher formants in the perception of vowels. *IEEE Transactions on Audio and Electroacoustics, AU-16*, 73-77.

Hawkins, S., & Stevens, K. N. (1985). Acoustic and perceptual correlates of the non-nasal - nasal distinction for vowels. *Journal of the Acoustical Society of America, 77*, 1560-1575.

House, A. S., & Stevens, K. N. (1956). Analog studies of the nasalization of vowels. *Journal of Speech and Hearing Disability, 21*, 218-232.

Houtgast, T. (1977). Auditory-filter characteristics derived from direct-masking data and pulsation-threshold data with a rippled noise masker. *Journal of the Acoustical Society of America, 62*, 409-415.

Joos, M. (1948). Acoustic phonetics (*Language Monograph 23*; Linguistic Society of America at Waverly Press, Baltimore, MD).

Klatt, D. H. (1981). Prediction of perceived phonetic distance from short-term spectra—a first step. *Journal of the Acoustical Society of America, 70*, Suppl. 1, S59.

Klatt, D. H. (1982). Speech processing strategies based on auditory models. In R. Carlson & B. Granström (Eds.), *The representation of speech in the peripheral auditory system* (pp. 181-196). New York: Elsevier.

Klatt, D. H. (1985). A shift in formant frequencies is not the same as a shift in the center of gravity of a multi-formant energy concentration. *Journal of the Acoustical Society of America, 77*, Suppl. 1, S7.

Miller, R. L. (1953). Auditory tests with synthetic vowels. *Journal of the Acoustical Society of America, 25*, 114-121.

Moore, B. C. J., & Glasberg, B. R. (1981). Auditory filter shapes derived in simultaneous and forward masking. *Journal of the Acoustical Society of America, 70*, 1003-1014.

Moore, B. C. J., & Glasberg, B. R. (1983). Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *Journal of the Acoustical Society of America, 74*, 750-753.

Mushnikov, V. N., & Chistovich, L. A. (1972). Method for the experimental investigation of the role of component loudnesses in the recognition of a vowel. *Soviet Physics-Acoustics, 17*, 339-344.

Ohala, J. J. (1974). Experimental historical phonology. In J. M. Anderson & C. Jones (Eds.), *Historical linguistics II: Theory and description in phonology* (pp. 353-389). Amsterdam: North-Holland.

Paliwal, K. K., Ainsworth, W. A., & Lindsay, D. (1983). A study of two-formant models for vowel identification. *Speech Communication, 2*, 295-303.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America, 24*, 175-184.

Ruhlen, M. (1978). Nasal vowels. In J. H. Greenberg, C. A. Ferguson, & E. A. Moravcsik (Eds.), *Universals of human language, Vol. II, Phonology* (pp. 203-242). Stanford: Stanford University Press.

Sachs, M. B., & Young, E. D. (1980). Effects of nonlinearities on speech encoding in the auditory nerve. *Journal of the Acoustical Society of America, 68*, 858-875.

Schourup, L. (1973). A cross-language study of vowel nasalization. *Ohio State Working Papers in Linguistics, 15*, 190-221.

Seneff, S. (1985). *Pitch and spectral analysis of speech based on an auditory synchrony model.* Cambridge: MIT Research Laboratory of Electronics, Technical Report 504.

Sidwell, A., & Summerfield, Q. (1985). The effect of enhanced spectral contrast on the internal representation of vowel-shaped noise. *Journal of the Acoustical Society of America, 78*, 495-506.

Stevens, K. N., Fant, G., & Hawkins, S. (1987). Some acoustical and perceptual correlates of nasal vowels. In R. Channon & L. Shockey (Eds.), *Festschrift for Ilse Lehiste* (pp. 241-254). Dordrecht, Holland: Foris.

Sundberg, J., & Gauffin, J. (1982). Amplitude of the voice source fundamental and intelligibility of superpitch vowels. In R. Carlson & B. Granström (Eds.), *The representation of speech in the peripheral auditory system* (pp. 223-228). New York: Elsevier.

Syrdal, A. K. (1985). Aspects of a model of the auditory representation of American English vowels. *Speech Communication, 4*, 121-135.

Syrdal, A. K., & Gopal, H. S. (1986). A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of the Acoustical Society of America, 79*, 1086-1100.

Traunmüller, H. (1981). Perceptual dimension of openness in vowels. *Journal of the Acoustical Society of America, 69*, 1465-1475.

Traunmüller, H. (1982). Perception of timbre: Evidence for spectral resolution bandwidth different from critical band? In R. Carlson & B. Granström (Eds.), *The representation of speech in the peripheral auditory system* (pp. 203-208). New York: Elsevier.

Verbrugge, R. R., Strange, W., Shankweiler, D. P., & Edman, J. R. (1976). What information enables a listener to map a talker's vowel space? *Journal of the Acoustical Society of America, 60*, 198-212.

Wright, J. T. (1986). The behavior of nasalized vowels the perceptual vowel space. In J. J. Ohala & J. J. Jaeger *Experimental phonology* (pp. 45-67). Orlando, FL: Academic.

FOOTNOTES

Journal of the Acoustical Society of America, 87(6), 2684-2704 (1990).

†Program in Linguistics, University of Michigan, Ann Arbor.

††Department of Linguistics, University of Cambridge, Cambridge, United Kingdom.

Haskins Laboratories Status Report on Speech Research
1991, SR-105/106, 215-230

On the Perception of Speech from Time-varying Acoustic Information: Contributions of Amplitude Variation*

Robert E. Remez† and Philip E. Rubin

The cyclic variation in the energy envelope of the speech signal results from the production of speech in syllables. This acoustic property is often identified as a source of information in the perception of syllable attributes, though spectral variation can also provide this information reliably. In the present study of the relative contributions of the energy and spectral envelopes in speech perception, we employed sinusoidal replicas of utterances, which permitted us to examine the roles of these acoustic properties in establishing or maintaining time-varying perceptual coherence. Three experiments were carried out to assess the independent perceptual effects of variation in sinusoidal amplitude and frequency, using sentence-length signals. In Experiment 1, we found that the fine grain of amplitude variation was not necessary for the perception of segmental and suprasegmental linguistic attributes; in Experiment 2, we found that amplitude variation was nonetheless effective in influencing syllable perception, and that in some circumstances it was crucial to segmental perception; in Experiment 3, we observed that coarse-grain amplitude variation, above all, proved to be extremely important in phonetic perception. We conclude that in perceiving sinusoidal replicas, the perceiver derives much from following the coherent pattern of frequency variation and gross signal energy, though probably derives rather little from tracking the precise details of the energy envelope. These findings encourage the view that the perceiver uses time-varying acoustic properties selectively in understanding speech.

In discussions of speech perception, the stability of the listener's phonetic impressions has often been contrasted with the variability of the underlying acoustic patterns (Fant, 1962; Fowler & Smith, 1986; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). These descriptive and theoretical accounts have been aimed at establishing the principles that govern the acoustic realization of phonetic segments intended by the talker, and the principles that in turn govern the perceptual realization of phonetic impressions by the listener. Though knowledge of the speech signal has developed greatly through this search for acoustic cues to phonetic identity, a satisfactory account of the listener's perceptual finesse eludes us. The difficulty may be traced to two critical facts about speech: There does not seem to be a core set of acoustic cues (Liberman & Cooper, 1972); and the variability of the signal structure does not appear to indicate a normal set of acoustic elements about which variation occurs (Bailey & Summerfield, 1980).

Although accounts of perception based on discrete acoustic cues and cue combinations retain a straightforward and durable appeal (Massaro, 1987; Stevens & Blumstein, 1981), some investigators have taken variability to be central in describing speech perception by the human listener, and they have therefore have pursued the hypothesis that speech perception, however it may be embodied, depends in part on coherent time-varying properties of the signal (e.g., Best, Studdert-Kennedy, Manuel, & Rubin-Spitz, 1989; Jenkins, Strange, & Edman, 1983; Kewley-Port & Luce, 1984). Our studies have likewise supported the claim that the listener can draw phonetic information from speech signal variation, in some

The authors gladly acknowledge the generous advice and stern encouragement of Jessica Lang, Paula Payton, Jan Rabinowitz, Amanda Steinberg, Ross Young, and the anonymous reviewers of an earlier submission of Experiment 1. This research was supported by grants from NIDCD (00308) to R. E. Remez, and from NICHHD (01994) to Haskins Laboratories.

sense independently of the specific acoustic elements composing the signal, and independently of specific auditory impressions of speech sounds (Remez & Rubin, 1983; Remez, Rubin, Pisoni, & Carrell, 1981). In this research, we have employed a sinusoidal replication technique, in which synthetic acoustic patterns are matched to gross spectral changes, but not to fine acoustic details, of a speech signal. Perceptual tests with these materials provide evidence of the role of time-varying information in speech perception.

In our studies of perceptual susceptibility to phonetic information carried by sinusoidal replicas of speech signals, three or four pure tones have been used to represent the frequency and amplitude variation of oral, nasal, and fricative formants of natural utterances. The result is a signal that differs drastically in its physical properties from natural speech. Typically, speech exhibits some consistent acoustic attributes: a pulse train and a harmonic series arising from the glottal excitation; broadband natural resonances of the supralaryngeal vocal tract, the formants; and aperiodic elements: release bursts, low-energy aspirate formants, and turbulence in fricative formants—to note three. Sinusoidal replicas of speech lack this fine grain of acoustic detail on which most characterizations of perceptual cues implicitly rely, yet they are comprehensible despite the absence of these familiar acoustic products of vocalization (Remez et al., 1981). (See Figure 1.) On the basis of studies employing sinusoidal replicas of natural utterances, and in agreement with related results found with more conventional acoustic techniques, we have learned that the perception of utterances may rely as much on the patterns composed by the elementary acoustic cues as on the specific psychoacoustic effects of the signal elements themselves.

Although our tests have suggested that the frequency variation of a sinusoidal sentence is perceptually effective, this interpretation can be challenged. Because each tonal component has represented both the frequency and amplitude variation of a natural resonance, a signal of this description delivers veridical formant amplitude information on anomalous carriers, which are several simultaneously varying sinusoids. In contrast to the predictions of our original hypothesis, which emphasized the coherent variation in tone frequency, we may alternatively find that tone amplitude variation plays a prominent perceptual role when sinusoidal signals are heard phonetically. In fact, a number of views of the perception of fluent natural speech propose a processing stage in which lexical patterns are hypothesized from metrical structure, and especially so when phonetic information is defective or ambiguous (see, e.g., Cutler & Foss, 1977; Cutler & Norris, 1988; Huggins, 1978; Nakatani & Schaffer, 1978). Perhaps a subject who attempts to transcribe a sinusoidal sentence employs such guesswork after exploiting the amplitude variation to perceive the sentence meter and rhythm.

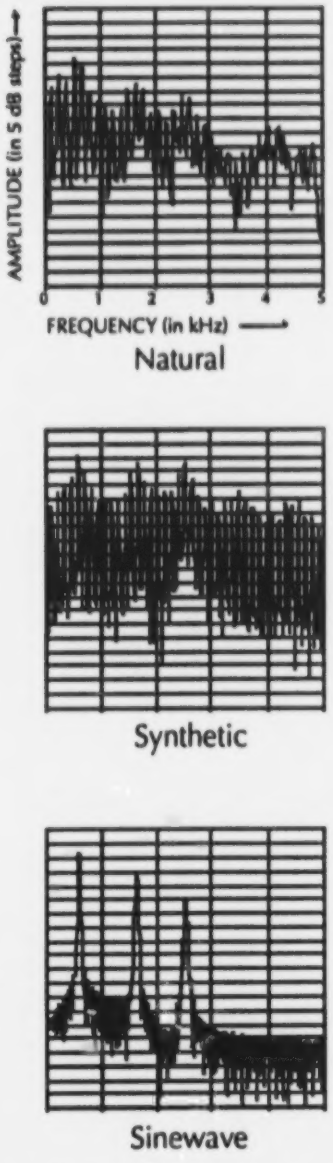


Figure 1. **The three panels present short-time spectra of three kinds of signal. Top: natural speech. Middle: cascade-type synthetic speech. Bottom: sinusoidal replication of speech.**

If this is true of the perception of sinusoidal sentences, then the listener does not perceive the message as we have claimed, from phonetic information preserved primarily in the time-varying pattern of tone frequencies. Moreover, this alternative to our frequency-based claim seems especially appealing, considering that the sensory derivation of this aspect of the signal, the envelope of a sinusoidal replica, probably occurs in the same way as it does for a natural signal, by summing the energy across spectral components. This encourages the possibility that the veridical amplitude envelope of a sinusoidal signal supplies phonetic information.

Accordingly, in the three experiments reported here, we attempted to resolve this issue by determining the perceptual effectiveness of tone amplitude variation in providing phonetic information within the context of sinusoidal replication. Experiment 1 demonstrated that the perceiver can do without this acoustically veridical property of sinusoidal sentence replicas under some circumstances, as our initial frequency-based hypothesis claimed; Experiment 2 solved part of the puzzle by establishing the sufficiency of tone frequency and amplitude variation as information in the perception of syllabic if not segmental properties; and Experiment 3 revealed that gross changes in the energy envelope, as opposed to the fine structure of the envelope, can have a great influence on the perception of speech from time-varying information.

EXPERIMENT 1

In the first study, we sought to determine the effect of veridical amplitude variation in the perception of sinusoidal sentences. The test called for two kinds of perceptual report from subjects, under four conditions of sinusoidal replication designed to identify the sufficiency of variation in sinusoidal frequency and amplitude.

The subjects were asked both to transcribe a sentence and to report the number of syllables that occurred. In the first condition, a sinusoidal pattern presented both the resonance frequency and the amplitude of each of the lowest three formants of a natural utterance. In the second condition, listeners heard a pattern that preserved the frequency variation of the first three formant centers, but it was presented at a constant level of energy throughout its duration. In the third condition, the sinusoidal pattern preserved the frequencies of the first three formant tracks, but with a misleading amplitude contour imposed on it. In the fourth condition, the sinusoidal pattern had the natural formant amplitude variation, but the tones in this case exhibited constant frequency, so that the effect of the natural amplitude contour could be estimated without concurrent frequency variation.

If the amplitude variations of the sinusoidal signal provided information about the prosodic structure of the utterance, and if the subjects relied primarily on this source of information in performing the two tasks that we set for them, then syllable counting would be accurate in the first and the fourth conditions, and poorer in the second and third conditions. If subjects perceived the phonetic sequence on the basis of the time-varying properties of frequency variation, however, transcription and syllable counting would be accurate in all conditions but the last, in which there was no frequency variation.

Method

Subjects

Fourteen adults with normal hearing in both ears made up each of the four groups that were tested, blocked by the four synthesis conditions. The subjects were drawn from sections of introductory psychology classes, and they received course credit for their participation. All were native speakers of English, and none had participated in any other experiments in which sinusoidal signals had been employed.

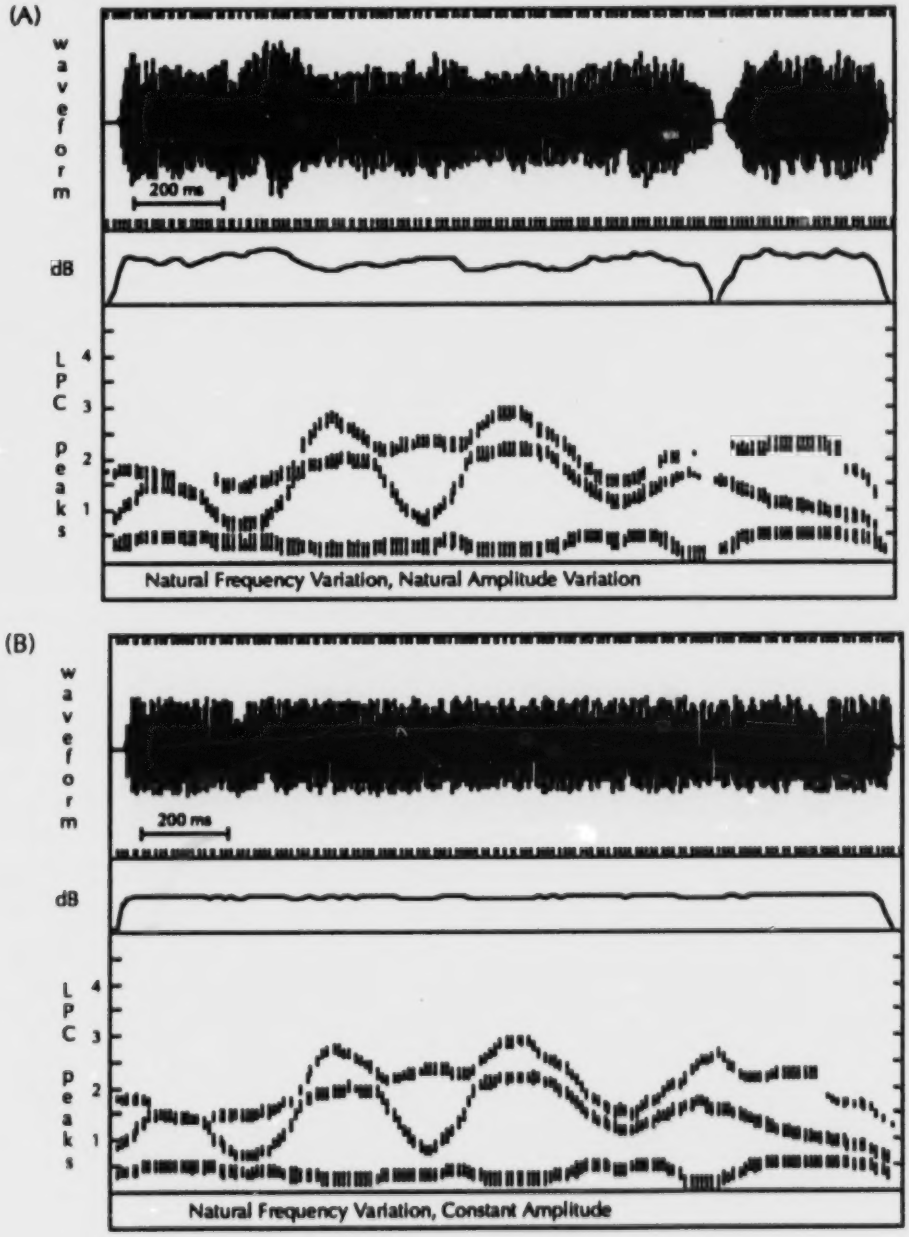
Acoustic Test Materials

Four different sinusoidal patterns produced by the sinewave synthesizer at Haskins Laboratories were used. This software synthesizer accepts frequency, amplitude, and duration values for setting the parameters of digital oscillators, and it calculates the resulting waveform with 12-bit amplitude resolution at a designated frame rate. A sentence ("Where were you a year ago?") uttered by one of the authors was used as the model for the four sinusoidal patterns.

The natural utterance was recorded on audiotape (Scotch No. 208) with a high-quality voice microphone (Shure SM87) in a small sound-attenuating chamber (IAC Model 400A), and then converted to digital records by filtering it (4.5-kHz low-pass, -40-dB/octave rolloff) and sampling it at 10 kHz, using a PCM system implemented on a DEC VAX-11/780. The sampled natural speech data were then analyzed with the technique of linear prediction, to determine the center frequency and amplitude of each of the lowest three formants. Formant frequencies and amplitudes were estimated at 10 ms intervals through the utterance, and these derived values

were then appropriated for use as a table of sinusoidal synthesis specifications.

The four patterns that we constructed are shown in Figure 2, and they correspond to the four perceptual tests of sinusoidal replication. Figure 2A shows the pattern used in the first test, which contained three sinusoids following the frequency and amplitude variations of the original natural utterance. Figure 2B shows the pattern tested in the second condition, which preserved the frequency variation of the spectrum, but with each tone at constant power throughout the duration of the sentence. The second tone had approximately half the power of the first, and the third had approximately half that of the second. Figure 2C shows the signal that was used in the third test, in which a misleading artificial amplitude pattern was imposed on each tone, though the frequency variations of each tone follow the values derived from the natural model. Last, Figure 2D shows the pattern of the fourth test, exhibiting natural amplitude values carried by constant-frequency tones, the lowest at 500 Hz, the second at 1500 Hz, and the third at 2500 Hz.



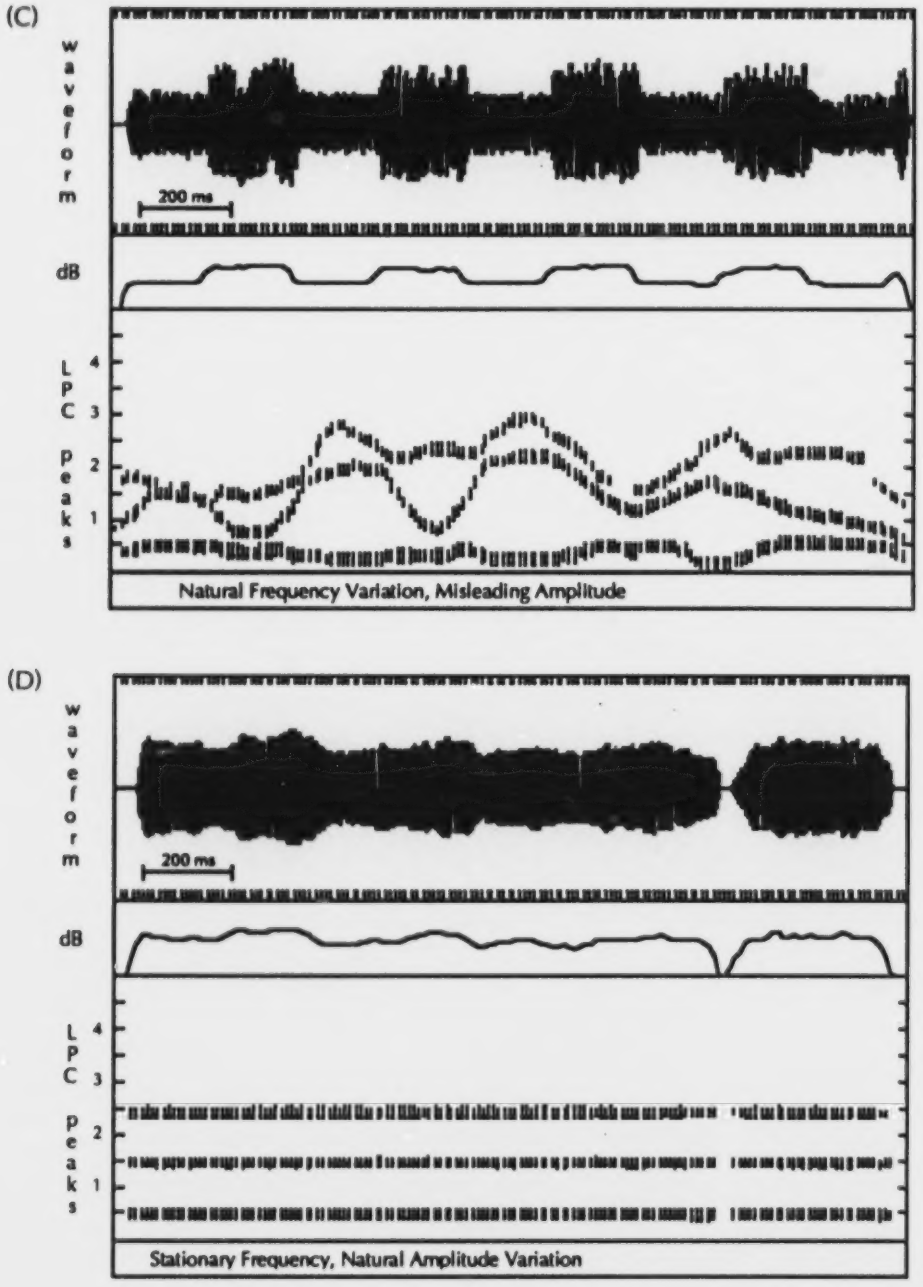


Figure 2. (A, opposite page) **The waveform (top), rms energy in decibels (middle), and spectral analysis for the natural amplitude test item (bottom).** *(B, opposite page)* **The waveform (top), rms energy in decibels (middle), and spectral analysis for the constant amplitude test item (bottom).** *(C)* **The waveform (top), rms energy in decibels (middle), and spectral analysis for the misleading amplitude test item (bottom).** *(D)* **The waveform (top), rms energy in decibels (middle), and spectral analysis for the test item possessing natural amplitude variation imposed on stationary-frequency tones (bottom).**

The synthetic sinusoidal patterns were converted from digital records to analog signals, recorded on half-track .25-in. audiotape (Scotch No. 208) at Haskins Laboratories, and were presented to subjects via tape playback using an Otari MX-5050B tape recorder and a Crown D-74 power amplifier. Listeners sat in carrels in a sound-shielded room, and signals were presented binaurally at an approximate level of 65 dB SPL over matched and calibrated Telephonics TDH-39 headsets.

Procedure

This experiment was composed of four tests, each with different listeners. Each test corresponded to one of the four sinusoidal signals shown in Figure 2: (1) natural frequency and amplitude variation; (2) natural frequency variation, with constant amplitude; (3) natural frequency variation, with artificial (and presumably) misleading amplitude variation; and (4) constant-frequency sinusoids with natural amplitude variation. A single sentence was presented four times in succession, separated by 1 sec, at the conclusion of which the subjects reported by writing their responses in prepared booklets.

Listeners were tested in groups of 6 or fewer. They were briefly instructed that synthetic speech was to be presented over the headphones, and they were asked to mark an answer sheet with (1) their impression of the number of syllables in the computer's utterance, and (2) a transcription of the sentence that the computer produced.

Results

The outcome of the tests, shown in Figure 3, was straightforward. The transcription test revealed that subjects were not hindered by defective amplitude properties of the signal as long as the information carried by tone frequency variation was available. Transcription performance for three tests was scored as the number of syllables correctly reported, with a maximum of seven if the sentence was transcribed completely. The group performance levels for the tests of the effect of amplitude envelope when tone frequency varied concurrently are shown in the top panel of Figure 3. (No transcriptions were scored for the fourth test in which tone frequency did not vary, and in which case the subjects did not report hearing phonetic segments.) These performance levels exceed a score of 0 syllables transcribed correctly, and they do not differ significantly from each other, as was shown in two statistical tests. First, the analysis of variance performed on these data revealed no significant effect of amplitude envelope on the accuracy of transcription [$F(2,36) = 0.64$, $p>.5$]. Second, the grand mean differed significantly from a score of zero [$t(38) = 11.19$, $p<.001$]. Because some subjects failed to follow the test instructions, we were unable to include the reports of 1 in the second test, and of 2 in the third.

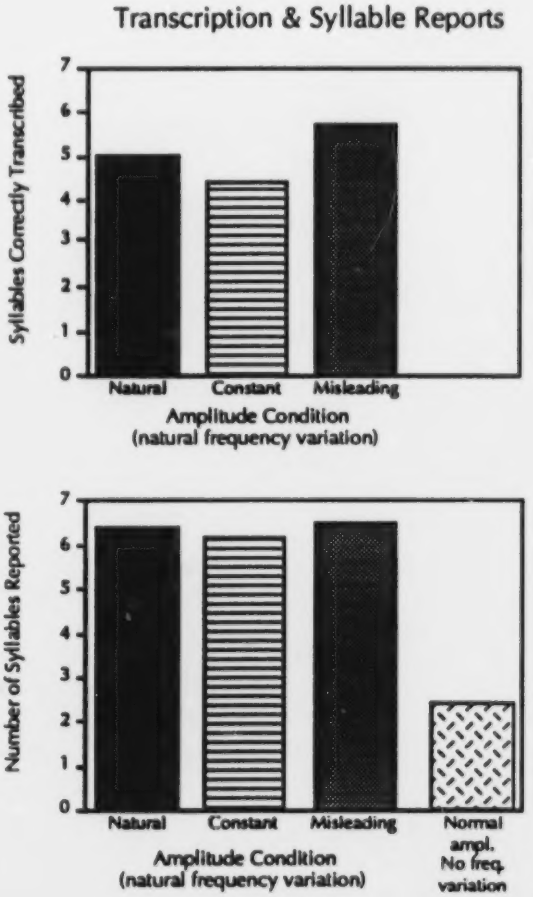


Figure 3. Group performance on the transcription task (top) and on the syllable counting task (bottom) of Experiment 1.

The results of the syllable counting test were surprising; they indicated that the perception of syllabic attributes of the sinusoidal utterance depended less on the amplitude envelope than on frequency variation. Group averages for reports of syllable counts in the four conditions are shown in the bottom panel of Figure 3. The analysis of variance performed on these data showed a significant effect of the acoustic conditions [$F(3,48) = 42.91$, $p<.001$], and a Scheffe post hoc means test confirmed the difference ($p<.001$) between the fourth test (stationary frequency, concurrent

natural amplitude variation) and the other three tests (natural frequency and amplitude variation, natural frequency variation at constant power, and natural frequency variation with misleading amplitude variation).

Discussion

On the basis of this first experiment, it seems that a sinusoidal replica preserves useful acoustic information for the phonetic composition of an utterance solely in its tone frequency variation. We may reject the hypothesis that the listener is only able to transcribe the phonetic properties of sinusoidal sentences by combining prosodic information carried in the veridical variation of the sinusoidal amplitude envelope with ambiguous or incomplete information supplied by the anomalous sinusoidal carriers. The evidence on this point provided by our listeners is compelling. They were unable to follow the syllable structure of the utterance in the crucial fourth condition, when the natural amplitude variation was presented without concurrent frequency variation. In that condition, the single salient linguistic property conveyed by the energy envelope was apparently the closure of the vocal tract during the production of a stop consonant, which may explain the consistent report that the pattern consisted of two syllables—one preclosure, one postclosure.

A minimal role of energy tracking in speech perception is also consistent with the results of the third condition, in which subjects were able to apprehend the phonetic detail even when the energy contour was grossly inappropriate to the segments subordinate to it. It seems that listeners who transcribed these sinusoidal replicas of speech must have relied on information about the phonetic sequence available primarily in the frequency variation, as we have claimed from the outcome of our prior research (Remez, 1987; Remez & Rubin, 1983, 1984; Remez, Rubin, Nygaard, & Howell, 1987; Remez et al., 1981).

Although this study revealed that listeners may disregard a rather consistent acoustic correlate of the syllable pattern, it does not permit us to draw a strong and general conclusion about phonetic perception from time-varying sinusoidal replicas, nor about ordinary speech perception. Though we have shown that veridical amplitude variation is not essential in perceiving tonal analogs of speech, we have not shown that amplitude variation is disregarded when it is available. Perhaps there are differences across utterances in the effectiveness of amplitude information, and Experiment 1 we may have employed an utterance in which the correspondences of energy envelope and syllable structure were less than straightforward. In any case, a fair test of amplitude as a source of information requires diversity in the linguistic properties of the test materials. The phone classes comprised by the sentence used in Experiment 1 were convenient but not representative of the inventory of English, consisting of liquid consonants, vowels, and a single stop consonant. Ordinary speech is not similarly restricted, and a general claim requires a less restricted test.

In Experiment 2, we therefore employed a sentence with nasal consonants, voiceless stops, voiced and voiceless fricatives, and consonant clusters, to remove this limitation on the first experiment. These segments are distinguished in the amplitude envelope to different degrees; stops and affricates especially are usually marked in the waveform with silence (Halle, Hughes, & Radley, 1957), and they may be viewed as the likeliest kind of phonetic segment to depend perceptually on information provided by amplitude. By adopting a factorial design, we also expected to clarify the relative independent contributions of frequency and amplitude variation to intelligibility and to judgments of syllable numerosity.

EXPERIMENT 2

The objective of Experiment 2 was to provide a more general test of the findings of the first experiment. Although it appeared that perceivers were indifferent to the only acoustically veridical component of a sinusoidal replica, our test had involved a sentence and a set of conditions that might have obscured the independent perceptual roles of tone amplitude variation and tone frequency variation. To provide the remedy, two aspects of the design were modified in Experiment 2. First, a different sentence was used, in order to replicate the acoustic treatments of Experiment 1 with greater phonetic and acoustic variety. Second, an extended set of conditions was used to examine cases of natural, constant, and misleading energy envelopes imposed on a set of natural, constant, and misleading frequency patterns, in a factorial design. This combination produced a test that was less phonetically restricted, in which the potential contributions of tone frequency and energy could be assessed. Again, we appraised the impact of these acoustic properties on perception of the phonetic and syllable pattern by collecting two reports on each trial: (1) a report of the number of syllables in the utterance, and (2) a transcription of the words.

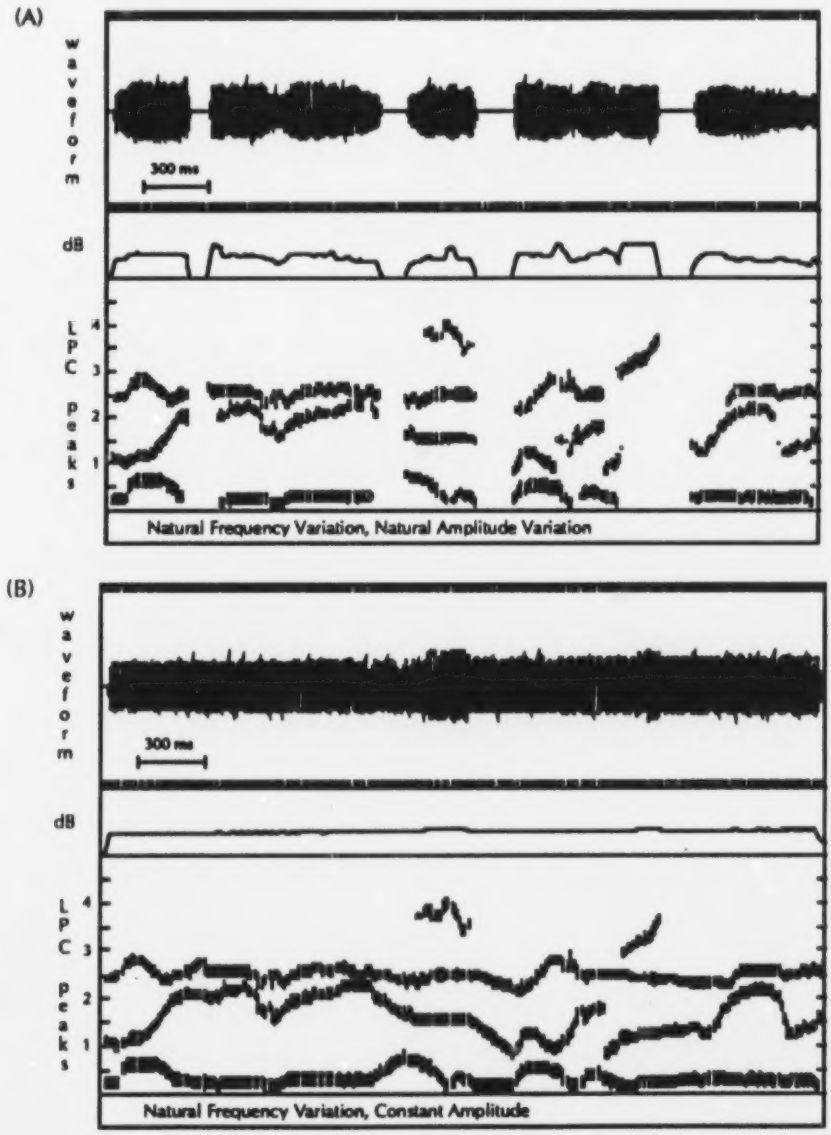
Method

Subjects

Two hundred and twenty-two subjects were tested overall, in nine test conditions. Some listeners were paid for participating, while others were drawn from the introductory psychology subject pool. Each reported normal hearing in both ears, and none had participated in any other experiments in which sinusoidal signals had been employed.

Acoustic test materials

Nine sinusoidal sentence patterns were created on the model of a natural utterance ("My t. v. has a twelve-inch screen.") produced by one of the authors. As in the case of Experiment 1, this utterance was recorded on tape, low-pass filtered at 4.5 kHz, and sampled by the VAX at 10 kHz with 12 bit resolution. The digital records of the utterance were analyzed at 10-msec intervals with the technique of linear prediction, to derive frequency and amplitude estimates of the oral, nasal, and fricative formants. After several erroneously estimated frequency values had been corrected, these analysis data were used to compose a table of sinusoidal synthesis specifications.[1] Incorporating both natural frequency and amplitude variation, this synthesis parameter set was used to make the *natural frequency, natural amplitude* test item, which is illustrated in Figure 4A.



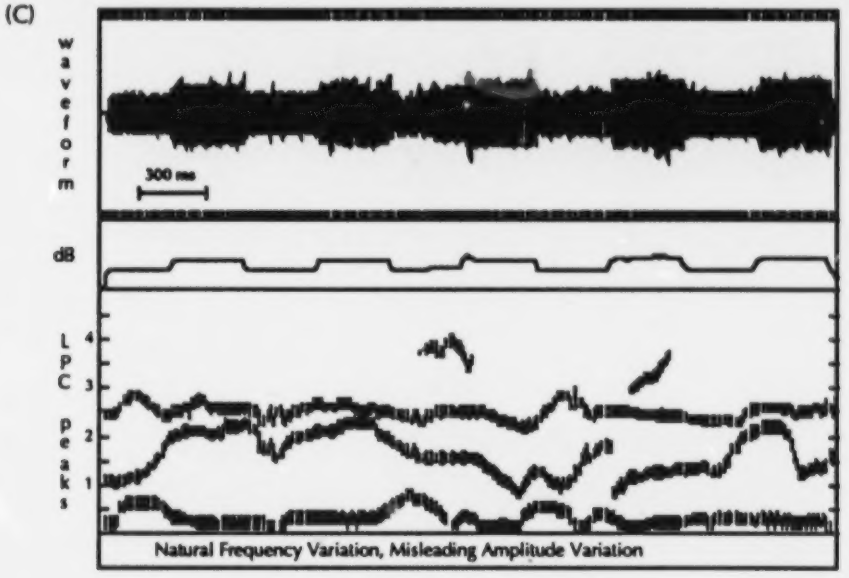


Figure 4. (A, opposite page) **Natural frequency, natural amplitude test item, sentence "My t.v. has a twelve-inch screen."** *(B, opposite page)* **Natural frequency, constant amplitude item.** *(C)* **Natural frequency, misleading amplitude test item.**

Eight variant patterns were then composed by modifying the synthesis parameters that had preserved the natural frequency and amplitude values. In all, three different kinds of frequency variation—the natural values, a misleading frequency pattern, and a stationary frequency pattern—were crossed with three kinds of amplitude variation—the natural values, a misleading amplitude pattern similar to that used in Experiment 1, and an unchanging amplitude pattern. The construction of each of the variants was straightforward:

Natural frequency, constant amplitude. To impose constant amplitude while preserving natural frequency variation, the amplitudes of Tones 1, 2, and 3 were each fixed throughout the utterance. A natural rolloff of -6 dB/octave was approximated across the three tones replicating the oral resonances. This manipulation also maintained the constant level through the portions of the sinusoidal replica that corresponded to stop closures. In addition, the intermittent and brief introduction of a fourth tone, correlated with the /z/ frication in [hæz] and the affricate-fricative /čs/ sequence in [mčskrin] was presented similarly by imposing a constant energy level during the duration of each brief tone. See Figure 4B for an illustration of this item.

Natural frequency, misleading amplitude. To impose an arbitrarily misleading energy pattern while preserving natural frequency variation, the synthetic pattern was divided conceptually into 310-msec portions, every other one of which was created with a 20% increase in tone amplitude relative to the instance of constant frequency. This produced 10 individual amplitude episodes, or five paired sequences. Again, the tones maintained the energy through portions of the utterance corresponding to stop closures. See Figure 4C for an illustration of this item.

Misleading frequency, natural amplitude. To compose a pattern of misleading tone frequency values, we imposed a sinusoidal modulation on each of four component tones, centered respectively at 500, 1500, 2500, and 3500 Hz, with a maximum excursion of ±250 Hz, and a period of modulation of 500 msec. This resulted in a four-tone frequency pattern that cycled approximately two and a half times over the duration of the signal. Again, this pattern of frequency values was combined with the formant amplitude values derived from the natural utterance, resulting in a sinusoidal pattern exhibiting misleading frequency variation and natural amplitude variation.

Misleading frequency, constant amplitude. The pattern of sinusoidally modulated tone frequencies was combined with constant-level amplitude values. The result was the pattern shown in Figure 5A, which also exhibited a natural rolloff.

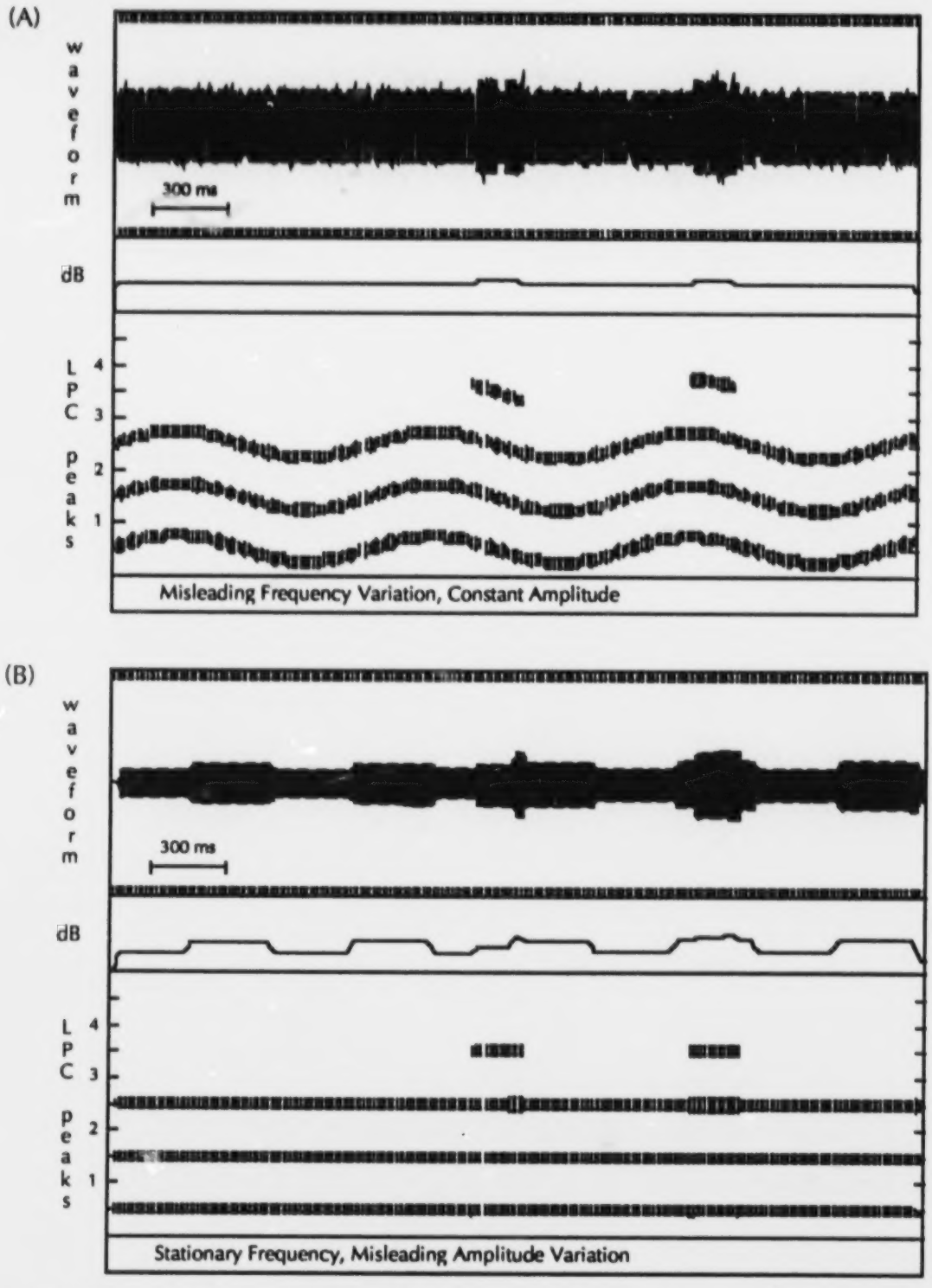


Figure 5. (A) **Misleading frequency, constant amplitude test item.** *(B)* **Stationary frequency, misleading amplitude test item.**

Misleading frequency, misleading amplitude. Amplitude portions differing in energy by 20% again alternated throughout the duration of this signal. This misleading amplitude pattern was combined with the sinusoidally modulated tone frequency values.

Constant frequency, natural amplitude. To remove natural frequency variation while preserving natural amplitude variation, the tone frequency values specified in the natural frequency, natural amplitude item were set to 500, 1500, and 2500 Hz, in the respective cases of Tones 1, 2, and 3, and to 3500 Hz for the brief intermittent "fricative" Tone 4.

Constant frequency, constant amplitude. The synthesis values eliminated all frequency and amplitude variation by combining stationary frequency values of 500, 1500, 2500, and 3500 Hz for the four component tones and constant amplitude levels. Again, the natural rolloff of -6 dB/octave was imposed across the tones.

Constant frequency, misleading amplitude. The amplitude pattern again contained alternating portions differing in energy by 20%, imposed on tone frequency values that were stationary throughout the duration of the signal. Figure 5B shows this test item.

The digital records of the synthetic sinusoidal patterns were converted to analog signals and recorded on audiotape. Listening tests were presented via tape playback. The subjects sat in carrels in a sound-shielded room, wearing matched and calibrated Telephonics TDH-39 headsets, and heard the signals binaurally attenuated to a level of approximately 65 dB SPL.

Procedure

There were nine conditions in this experiment, corresponding to the nine sinusoidal patterns varying in frequency and amplitude properties. The subjects were assigned randomly to conditions, and were tested in groups of 6 or fewer. As was the case in Experiment 1, subjects were instructed that synthetic speech was to be presented over the headphones, and were asked to report both an impression of the number of syllables in the computer's utterance, and also to transcribe the sentence that the computer produced.

Each session began with an eight-sentence pretest that was used both to promote the subject's susceptibility to sinusoidally replicated utterances, and, later, to assess that susceptibility. The sentences were drawn from the set of Egan (1948). At the immediate conclusion of the pretest, one of the nine test sentences was presented. Every sentence was presented eight times in succession, separated by 3 sec, with 10 sec between trials. The subjects reported their impressions by writing in prepared booklets.

Results

Pretest

The results of the eight-sentence pretest were used to determine whether the subjects were susceptible to the phonetic attributes of tone analogs of speech. In the present case, we found that 18% of the subjects (42 of them) failed to transcribe any sentences on the pretest, and they were therefore eliminated from the data set. This left 20 subjects in each of nine test conditions who passed the pretest and who presumably were sensitive to the phonetic effects of sinusoidal replication of speech.

Transcription performance

The statistical analysis of transcription reports was unnecessary, because there was only one condition in the set of nine in which transcription occurred: the combination of natural frequency and natural amplitude. With that congruence of acoustic properties, the average number of syllables reported correctly was 4.7. In neither of the two other conditions in which natural frequency values were presented, nor in the six conditions in which misleading or unchanging frequency values were presented, did average transcription performance exceed 1 syllable.

Syllable numerosity reports

The results of the tests of acoustic influence on judgments of syllable numerosity were determined by a two-way analysis of variance, crossing three frequency categories with three amplitude categories. The group means of apparent syllable reports are shown in Figure 6, in each of the nine conditions. Both main effects were observed: for frequency [$F(2,171)=15.88$, $p<.001$] and for amplitude [$F(2,171)=8.50$, $p<.001$]. Moreover, the interaction of these factors was also significant [$F(4,171)=7.86$, $p<.001$], as is apparent from the figure.

Post hoc tests (Newman-Keuls method of multiple comparisons, $\alpha=.05$) were used to assess the statistical differences among the nine means. To summarize the findings: First, accurate reports of syllable numerosity (eight syllables) were obtained in five conditions. Of course, this occurred in the case of natural frequency and amplitude presentation, but it was observed as well in the two cases of unnatural amplitude variation imposed on natural frequency variation,

and in the complementary cases of unnatural frequency variation exhibiting a natural energy envelope. This shows that both frequency variation and amplitude variation can provide sufficient information for some aspects of syllable perception. Second, in the extreme case of constant tone frequency within an unchanging energy envelope, the syllable reports approached a mean of 2, perhaps suggesting that the intermittent tone following the two momentary occurrences of a fricative formant was the residual property of the displays on which listeners based their perceptual reports. Last, in the three conditions in which either amplitude or frequency properties, or both, were misleading, the reports were statistically identical, approaching a mean of 6 syllables.

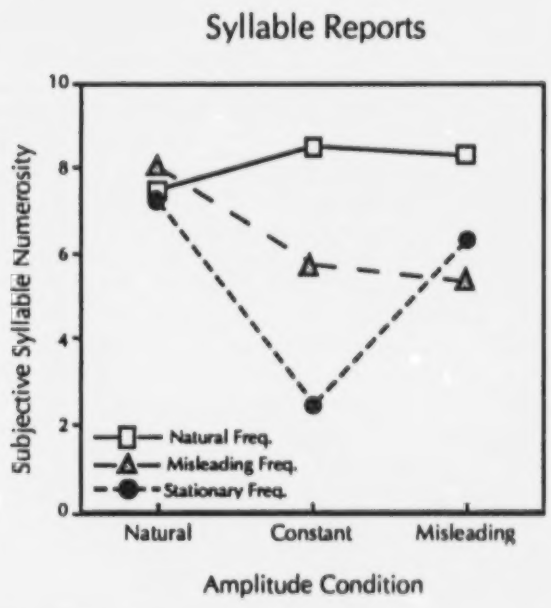


Figure 6. Average syllable numerosity reports in Experiment 2.

Discussion

Experiment 2 can clarify the issues pertaining to the perception of tone analogs if we consider three aspects of the outcomes: (1) the parallels between the findings here and those in Experiment 1; (2) the circumstances in which frequency and amplitude properties influenced judgments of syllable numerosity; and (3) the surprisingly poor transcription performance that was observed in two of the three natural frequency conditions. Collectively, these points supplement the conclusion suggested by the first experiment, by suggesting a role of both frequency and amplitude variation in the perception of sinusoidal sentences.

First, in the four conditions that the two experiments shared, the pattern of syllable numerosity results in Experiment 2 was similar to the findings in Experiment 1. Briefly, in both experiments, there was little effect of misleading or constant amplitude properties on the apparent syllable pattern when the natural frequencies were also presented. Because this was observed here in a sentence of rather diverse phonetic composition, we can be relatively confident that the results of Experiment 1 were not attributable to the preponderance of continuant consonants used in that test. Again, it seems that sufficient acoustic information relevant to the perception of the pattern of syllables in a sinewave sentence is to be found within the properties of frequency variation of the component tones.

Second, this hypothesis must be supplemented to accommodate the obvious influence of amplitude patterns that were revealed in test conditions featuring natural amplitude values imposed on tones with misleading or constant-frequency properties. The availability of the natural amplitude pattern was adequate to guarantee accurate syllable reports despite the elimination of natural frequency variation, suggesting that natural amplitude properties, much like natural frequency patterns, are perceptually salient and adequate. This finding clearly supplements the conclusion suggested by Experiment 1, which emphasized solely the adequacy of information borne by frequency variation. Furthermore, the three conditions in which tone frequencies did not vary show the unmistakable perceptual effects of the amplitude manipulations on perception. Altogether, then, this test established the sufficiency of amplitude variation and replicated the proof of the adequacy of frequency variation in the perception of some attributes of syllables. A set of parametric evaluations of the perceptual sensitivity to natural signal variation may reveal the specific physical properties of the signal or the auditory mechanisms permitting natural variation to block the effects of distorted or unnatural forms of change. The present results show that such study is clearly warranted.

Third, a striking discrepancy is to be found in considering the transcription results of Experiments 1 and 2, in contrast to the close agreement in syllable counting. Unlike what occurred in the outcome of Experiment 1, here

transcription performance deteriorated completely when anomalous amplitude patterns were used with natural frequency values. Because syllable reports were not similarly disrupted, this finding suggests that the phonetic impressions available to subjects in such conditions were prosodically related to the natural cases, but were unintelligible due to the obscurity of key phonetic features. Is this finding due to the imposition of unnatural amplitude contours that, by requiring tonal patterns to persist throughout the sentence, obscured the stop closures?

This speculation is encouraged by two kinds of evidence. The first, is the precedent of Experiment 1, in which transcription performance was unaffected by the amplitude pattern imposed on the sinusoidal components. Perhaps this immunity from the impaired performance that was observed in Experiment 2 was due solely to the absence of stop consonants from that sentence. Certainly, one highly restricted hypothesis that may be based on the two experiments is that amplitude may vary freely without perceptual consequence if it is imposed on frequency variation that replicates the spectral properties of speech, and if there are few or no stop consonants in the sentence. The test of this hypothesis requires a replication in which unnatural amplitude variation occurs while the silences associated with stop consonants are conserved. The second kind of evidence encouraging this speculation is found in the general findings of sinewave replication. In contrast with the emphasis of some perceptual research that has highlighted the role of acoustic details in phoneme recognition—brief elementary cues, in other words—sinewave studies suggest that the perceiver need not be a meticulous listener attentive to a panoply of brief acoustic elements. The evidence of sinusoidal replication reveals the operation of a different kind of sensory susceptibility for which the coarse grain of the acoustic signal is effective. While it remains to be demonstrated conclusively that the sensitivity to time-varying signal attributes plays a crucial role in more acoustically ordinary cases (though, see Whalen, 1984, for a potential instance), one principle tentatively urged by this work is that perception of speech may rely much more on the coarse acoustic grain in frequency and amplitude than previous views have allowed. The crucial test of this requires the specification of the grain size of perceptual information in speech.

In the cases of amplitude variation in Experiment 2, we find that the silence during stop closures was a kind of gross property—the *complete* absence of energy—that may bear significant perceptual information. To evaluate this proposal, we conducted a third experiment to determine whether transcription performance was restored by the presence of appropriate silences in the signal, regardless of other attributes of the energy envelope that may be anomalous. This test called only for three amplitude conditions incorporating natural frequency variation, though unlike the second experiment, silences reflecting stop closures were preserved in misleading and constant amplitude test items. Our conclusion favoring gross energy changes would be warranted if the introduction of silences permitted the perception of segmental phonetic properties despite fine-grain anomalies in the energy envelope.

EXPERIMENT 3

In Experiment 2, we saw that the variations in the energy envelope had the greatest effect on perception of syllables when frequency variation was unnatural and perhaps neutralized as a source of phonetic information. In contrast with the effects on syllable reports, the effects on transcription—a fairly direct measure of segment perception—were most dramatic in the conditions presenting natural frequency variation. It was surprising, nonetheless, to see transcribability deteriorate so completely in the two energy conditions that departed from the natural values. Experiment 1 offered no clue that this might occur, perhaps because the phonetic variety and composition of the test sentence in Experiment 1 made it immune to any such effect. To resolve the difference between these two findings—one suggesting that an anomalous energy envelope does not affect transcription, the other that it hugely affects transcription—required a third test.

In Experiment 3 we evaluated the hypothesis that phonetically important information can be conveyed by gross changes in the energy envelope. Though this hypothesis is newly invoked for explaining the perception of spoken messages from sinusoidal vehicles, it is an established finding in linguistic phonetics (Halle et al., 1957; cf. Fitch, Halwes, Erickson, & Liberman, 1980). Applied to the case at hand, the hypothesis predicts that the sentence used in Experiment 2, which had stop consonants in crucial positions, was difficult to perceive because the silences correlated with vocal tract closure were eliminated.[2] It also permits the conclusion, based on the findings in Experiment 1, that an anomalous amplitude envelope may not disrupt

segment perception when there are few or no stop consonants at stake.

Method

Subjects

Fifty-two subjects were tested. Each listener reported normal hearing in both ears, and none had participated in studies in which sinewave replicas of utterances had been employed. Some subjects received course credit, and others were paid for participating.

Acoustic test materials

Three sinusoidal sentence patterns based on the utterance "My t.v. has a twelve-inch screen" were used to compose the test. One, which had natural frequency and amplitude parameters, was identical to the pattern used in Experiment 2. The second and third patterns were derived, respectively, from the *natural frequency, misleading amplitude* test item and the *natural frequency, constant amplitude* test item used in Experiment 2. In both cases, the amplitude patterns that were imposed on the natural frequency variation remained misleading or constant, with the exception that silences occurring during consonant closures were restored by making the necessary changes to the sinewave synthesis parameters. Silence durations were adopted from the natural amplitude values.

Sinusoidal synthesis produced digital records of the waveforms, which were converted to analog signals and recorded on audiotape. Test signals were presented from tape, attenuated to approximately 65 dB SPL, over matched and calibrated headsets.

Procedure

A session began with an eight-sentence pretest, in the same manner as in Experiment 2, at the conclusion of which the critical test sentence was presented. Each sentence occurred eight times, separated by 3 sec, with 10 sec between trials. Subjects were asked to transcribe a sentence synthesized by a computer. There were three test conditions in this experiment, corresponding to the three different amplitude patterns imposed on the natural frequency values: (1) natural amplitude, (2) misleading amplitude with silent closures, and (3) constant amplitude with silent closures. Subjects were assigned randomly to conditions.

Results and Discussion

Seven subjects misunderstood the test instructions or transcribed none of the pretest sentences and were eliminated from the data set (7 of 52 = 13%), leaving 15 subjects in each of the three test conditions. The outcome of the test was quite clear: Transcription performance was good in all three conditions. Mean performance for the natural amplitude group was 5.2 syllables correct; for the misleading amplitude group it was 3.3 syllables; for the constant amplitude group it was 4.1 syllables. Two statistical tests determined that transcriptions differed from a hypothetical mean of 0 syllables correctly transcribed, but that performance did not differ across the three test conditions. First, the statistical difference between the grand mean of the data set and a hypothetical mean performance of 0 syllables correct was assessed by t test, which indicated a highly significant nonzero performance in these conditions ($t(44) = 10.42$, $p<.0005$). Second, a one-way analysis of variance performed on these data revealed no significant differences in performance among the three test conditions [$F(2,42)=1.896$, $p>.1$].

The conclusion prompted by these results fits well with those of our prior tests, pointing clearly to the phonetic perceptual importance of coarse-grain variation in signal amplitude. It resolves the discrepancy between the first and second experiments of this set, namely, in identifying the role of silence as a conspicuous acoustic marker of consonantal closures. By hindsight, we can see that the elimination of silent portions of the signal can only have had minimal effect in Experiment 1, inasmuch as the only segment affected was the single stop consonant in the test sentence. In Experiment 2, the deterioration of transcription performance was more pronounced due to the presence of stops throughout the sentence. Perceivers in the present test tolerated the departures from natural amplitude variation in fine detail, we may presume, because in coarse grain the perceptually critical closures were well evident.

GENERAL DISCUSSION

This set of three experiments was motivated by the need to assess the contribution of the sole acoustic aspect of sinusoidal sentence replicas that presents the listener a veridical aspect of the speech signal: the energy envelope. Sinewave replicas preserve the variation in amplitude of the individual vocal resonances, despite the fact that the rise and fall of the energy pattern is imposed on unnatural carriers. Although it had seemed exceedingly likely that frequency variation of the individual tones was the principal source of

information in the perception of sentence analogs, the empirical support for this hypothesis was equivocal. In fact, the role of metrical properties in lexical information processing was well supported, and one plausible acoustic basis for this metrical information is the cyclical variation in the energy envelope, the same veridical attribute of speech preserved by sinewave sentences. An alternative to our initial hypothesis—that the listener attends to time-varying frequency information, instead of momentary, elementary speech cues—was therefore warranted. In contrast, we considered the hypothesis that the listener is unable to derive much from tone frequency variation after all, but is able to transcribe such signals by relying on the veridical properties of the energy envelope when the weird timbre, or sharp spectral peaks, or harmonically unrelated components of tone complexes yield ambiguous or defective phonetic information.

In fact, the picture drawn by the three experiments reported here differs from both of these characterizations. There is no denying now that both frequency and amplitude variation are capable of producing impressions of syllable variation, independently of each other in some circumstances. More germane to the case of sinusoidal replicas, though, is the finding that phonetic perception seems to depend crucially on the concurrent availability of natural frequency variation and gross amplitude variation. As far as we can tell from immediate evidence, the perceptually critical amplitude information was, grossly, the presence or absence of signal, a correlate of obstruction of the vocal tract. Other less extreme variations in amplitude level had no measurable effect in our tests, and on that evidence they are consigned to the class of the fine acoustic grain. Converging tests are still required to evaluate this, to be sure, but it seems after all that the perceptually important aspect of the amplitude variation is to mark the stop consonants.

What does this result say about ordinary speech perception? Considered broadly, this result is understandable from the perspective of studies on the robustness of spoken communication. It appears that little information may be carried by the amplitude pattern alone, for amplitude distortion is rarely devastating to segment intelligibility (Klatt, 1985; Licklider, 1946; Miller, 1946). Moreover, despite the temptation to treat the syllable as if it were an acoustic unit—because the energy envelope is correlated with the cyclical opening and closing of the vocal tract—the syllable is also a linguistic unit. To the cross-language evidence on the composition of the syllable (e.g., that of Price, 1980) and to the phonological arguments (e.g., those of Kiparsky, 1979) we can add this perceptual evidence about the specific value of amplitude variation in signaling stops rather than syllable trains.

Many of the parallels between the listener's perceptual treatment of natural signals and sinusoidal replication remain to be shown. However, the agreement between our data and those in other relevant studies of language encourage the generalization of our present finding—that syllable perception depended on the same acoustic properties as did segment perception—to the claim that the syllable is a multiply determined linguistic unit corresponding to no simple property of the acoustic signal. Collectively, the diverse evidence suggests that perception of sinusoidally recoded signals is similar to the perception of natural speech, and this apparent congruence seems again to indicate that such tonal replicas preserve information ordinarily available in natural acoustic signals. The listener who contends with a sinusoidal sentence, in our view, makes use of this informative time-varying residue, and does little that is exceptional to perceive the linguistic properties. Nonetheless, in order to reveal the operation of perception from coarse-grain properties, it is essential to employ acoustic signals that leave the perceiver no alternative but to rely on nonelemental time-varying attributes.

To summarize, in three experiments, we have added to the evidence that the perception of sinusoidal replicas of speech signals is based on the coherent frequency variation of the tonal components. The first experiment in this report rules out the possible counterargument that perception occurs only through a process of attention to syllable properties first, as if this were possible on the acoustic basis of the veridical amplitude variation that the tones present. On the contrary, amplitude variation can count for little, and frequency variation for much. The second experiment showed that the perception of some attributes of syllables endured the elimination or distortion of natural frequency or energy variation. Last, the third experiment showed that the intelligibility of sinewave replicas depended on the concurrent availability of variation in frequency and in amplitude, though the fine grain of amplitude variation appeared to be negligible, perceptually, when the overall coarse properties were preserved. These findings permit us to extend the alternative to the familiar

characterizations of the perception of speech based on discrete cues—that perception can be keyed to acoustic variation, independently of the specific fine-grain acoustic details that compose the signal—which we have derived from considering the phenomenon of tone analogs of speech.

REFERENCES

Bailey, P. J., & Summerfield, Q. (1980). Information in speech: observations on the perception of [s]-stop clusters. *Journal of Experimental Psychology: Human Perception & Performance, 6*, 536-563.

Best, C. T., Studdert-Kennedy, M., Manuel S., & Rubin-Spitz, J. (1989). Discovering phonetic coherence in acoustic patterns. *Perception & Psychophysics, 45*, 237-250.

Cutler, A., & Foss, D. J.(1977). On the role of sentence processing. *Language & Speech, 20*, 1-10.

Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception & Performance, 14*, 113-121.

Darwin, C. J., & Bethell-Fox, C. E. (1977). Pitch continuity and speech source attribution. *Journal of Experimental Psychology: Human Perception & Performance, 3*, 665-672.

Egan, J. (1948). Articulation testing methods. *Laryngoscope, 58*, 955-991.

Fant, C. G. M. (1962). Descriptive analysis of the acoustic aspects of speech. *Logos, 5*, 3-17.

Fitch, H. L., Halwes, T., Erickson, D. M., & Liberman, A. M. (1980). Perceptual equivalence of two acoustic cues for stop-consonant manner. *Perception & Psychophysics, 27*, 343-350.

Fowler, C. A., & Smith, M. R.(1986). Speech perception as "vector analysis": An approach to the problems of invariance and segmentation. In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 123-138) Hillsdale, NJ: Erlbaum.

Halle, M., Hughes, G. W., & Radley, J.-P. A. (1957). Acoustic properties of stop consonants. *Journal of the Acoustical Society of America, 29*, 107-116.

Huggins, A. W. F. (1978). Speech timing and intelligibility. In J. Requin (Ed.), *Attention and Performance VII* (pp. 279-297). Hillsdale, NJ: Erlbaum.

Jenkins, J. J., Strange, W., & Edman, T. R. (1983). Identification of vowels in "vowelless" syllables. *Perception & Psychophysics, 34*, 441-450.

Kewley-Port, D., & Luce, P. A. (1984). Time-varying features of initial stop consonants in auditory running spectra: A first report. *Perception & Psychophysics, 35*, 353-360.

Kiparsky, P. (1979). Metrical structure assignment is cyclic. *Linguistic Inquiry, 10*, 421-441.

Klatt, D. H. (1985). A shift in formant frequencies is not the same as a shift in the center of gravity of a multiformant energy concentration. *Journal of the Acoustical Society of America, 77*, S7.

Liberman, A. M., & Cooper, F. S. (1972). In search of the acoustic cues. In A. Valdman (Ed.), *Papers in Linguistics and Phonetics to the Memory of Pierre Delattre* (pp. 329-338). The Hague: Mouton.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74*, 421-461.

Licklider, J. C. R. (1946). Effects of amplitude distortion upon the intelligibility of speech. *Journal of the Acoustical Society of America, 18*, 429-434.

Massaro, D. W. (1987). Categorical partition: A fuzzy logical model of categorization behavior. In S. Harnad (Ed.), *Categorical perception* (pp. 254-283). New York: Cambridge University Press.

Miller, G. A. (1946). Intelligibility of speech: Effects of distortion. In *Transmission and reception of sounds under combat conditions* (pp. 86-108). Washington DC: National Defense Research Committee.

Nakatani, L. H., & Schaffer, J. A.(1978). Hearing "words" without words: Prosodic cues for word perception. *Journal of the Acoustical Society of America, 63*, 234-245.

Price, P. J. (1980). Sonority and syllabicity: Acoustic correlate of perception. *Phonetica, 37*, 327-343.

Remez, R. E. (1987). Units of organization and analysis in the perception of speech. In M. E. H. Schouten (Ed.), *Psychophysics of speech perception* (pp. 419-432). Dordrecht: Martinus Nijhoff.

Remez, R. E., & Rubin, P. E. (1983). The stream of speech. *Scandinavian Journal of Psychology, 24*, 63-66.

Remez, R. E., & Rubin, P. E. (1984). On the perception of intonation in sinusoidal sentences. *Perception & Psychophysics, 35*, 429-440.

Remez, R. E., Rubin, P. E., Nygaard, L. C., & Howell, W. A. (1987). Perceptual normalization of vowels produced by sinusoidal voices. *Journal of Experimental Psychology: Human Perception & Performance, 13*, 40-61.

Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science, 212*, 947-950.

Stevens, K. N., & Blumstein, S. E. (1981). The search for invariant acoustic correlates of phonetic features. In P. D. Eimas & J. L. Miller (Eds.), *Perspectives in the study of speech* (pp. 1-38). Hillsdale, NJ: Lawrence Erlbaum Associates.

Whalen, D. H. (1984). Subcategorical phonetic mismatches slow phonetic judgments. *Perception & Psychophysics, 35*, 49-64.

FOOTNOTES

Perception & Psychophysics, 48(4), 313-325 (1990).

†Department of Psychology, Barnard College, Columbia University

[1]It is typical for linear prediction estimates of vocal resonances to contain frequency values that depart dramatically from the acoustic spectrum at points of rapid spectrum change. Such results are common at the onsets and offsets of silence associated with closures and releases of the vocal tract, as occur in instances of stop consonants. In order to correct the results of linear prediction analyses for use as synthesis parameters, it is therefore necessary to check the computed values against unbiased acoustic analyses, such as, for example, those produced by the sound spectrograph.

[2]An apparent stop closure without associated acoustic silence was observed by Darwin and Bethell-Fox (1977). In that instance, though, the impressions of the stop closure were also associated with a gross spectral change. But it was a gross change in the fundamental frequency of phonation, creating an impression of two talkers uttering speech in succession, rather than a silent closure within a single speech signal creating an impression of a closure within a single utterance.

Haskins Laboratories Status Report on Speech Research
1991, SR-105/106, 231-242

Subject Definition and Selection Criteria for Stuttering Research in Adult Subjects*

Peter J. Alfonso†

The principal topics assigned to this author were to review subject definition and selection criteria reported in experiments investigating speech behaviors in adult stutterers, to determine whether the typical description of these criteria is sufficient to meet the demands of scientific investigation and replication, and to discuss a number of topics that should be considered in the development of subject definition and selection criteria. Although the organization of this paper is largely motivated by the assigned topics, I have deviated somewhat by reviewing in Section I the definitions of stutterers and stuttering that typically appear in the literature. The review is followed by a discussion of how the precise and careful use of these definitions bear directly on the development of subject definition and selection criteria. Section II represents the primary focus of the paper and includes: (a) a review of typical subject definition and selection criteria, and (b) a discussion of a number of topics that should be considered in the development of these criteria.

The argument that appropriate subject definition and selection criteria is an essential component of good experimental technique is based on the ubiquitous observations that so many of the overt characteristics of the disorder are highly variable across subjects at all levels of measurement. The argument developed here is that until the many facets of the heterogeneity of stuttering are better understood, criteria should err on the side of over-defining rather than under-defining essential details about the stutterer and his or her behaviors.

Preparation of this paper supported in part by NIH NS-13870 and NS-13617 awarded to Haskins Laboratories. The author wishes to thank three anonymous reviewers for their comments.

Finally, because this paper centers on experiments dealing with adults, much of what is written here presumes that the procedures employed in the experiments of interest are common to that population and are more physiologically based and invasive than those commonly used with children who stutter. Problems associated with the development of subject definition and selection criteria in the latter population group are considered in detail in a paper entitled "Childhood stuttering: What is it and who does it?" by Edward Conture (1990).

SECTION I: DEFINING STUTTERING AND STUTTERERS

I.A. An overview

Definitions of stuttering continue to evolve with our theories and abilities to measure various aspects of the disorder. Traditionally, most definitions are descriptions of behaviors. They are typically presented as a comprehensive list of behaviors that are common to all stutterers and that differentiate stuttering from normal speech. An often cited example of a descriptive definition is given by Wingate: "The term 'stuttering' means: I. (a) Disruption in the fluency of verbal expression, which is (b) characterized by involuntary, audible or silent, repetitions or prolongations in the utterance of short speech elements, namely: sounds, syllables, and words of one syllable... II. Sometimes the disruptions are (e) accompanied by accessory activities involving the speech apparatus, related or unrelated body structures, or stereotyped speech utterances... III. Also, there are not infrequently (f) indications or report of the presence of an emotional state, ranging from a general condition of 'excitement' or 'tension' to more specific emotions... (g) The immediate source of stuttering is some incoordination expressed in

the peripheral speech mechanism; the ultimate cause is presently unknown and may be complex or compound" (Wingate, 1964). Most definitions include at least the following three descriptions of the verbal behavior associated with stuttering; involuntary, repetitions, and prolongations. For example, stuttering is defined in the International Classification of Diseases as "disorders in the rhythm of speech, in which the individual knows precisely what he wishes to say, but at the time is unable to say it because of an involuntary, repetitive prolongation or cessation of a sound" (World Health Organization, 1977, p. 202). For a good discussion of the various categories of definitions see Van Riper, 1982, Chapter 2.

I.B. Needs and future directions

I.B.1. Disfluency, dysfluency, and fluency. The intent prescribed to the terms "disfluency," "dysfluency," and "fluency" varies considerably as a function of the distinctions among theoretical models of stuttering. This issue is considered in sufficient detail elsewhere, for example, Ham (1989), Perkins (1984), and Wingate (1984a.b, 1988). There should be no disagreement, however, about the critical necessity to make explicit, especially for research purposes, the descriptions of the speech behavior(s) under observation. By way of example, I will throughout this paper retain distinct definitions for the terms "disfluency," "dysfluency," and "fluency," largely following the rationale reported by Wingate (1984a,b). *Disfluency* is used here as a general referent, pertaining to the usual and normal disruptions in the patterns of speech movements that are perceived as "fluent speech." *Dysfluency*, on the other hand, is used to mean abnormal disruption in the normal patterns of speech movements. *Fluency* is used in the perceptual sense, to mean the realization of flowing, smooth, and easily produced speech; that is, as an abstraction of the underlying articulatory gestures. Thus, a sample of speech produced by a normal talker that has been judged to be fluent, as defined here, may include, at the level of speech production, disfluent segments but not dysfluent ones. Once again, the critical point is that these terms, and the criteria developed to operationally define them, must be made explicit. It will be shown in Section II that the failure to do so can led to conflicting conclusions about a number of aspects related to stuttering.

I.B.2. Stutterer's self-identification and stuttering-identification. In defining stuttering and stutterers, infrequent attention seems to be given to the stutterer's "self-identification" and "stuttering-identification." In particular, the stutterer's identification of a dysfluency can be very important since, as discussed below in more detail, it is often the case that disagreement will occur in fluent-dysfluent judgments that are based on data representing different accessible levels of measurement, that is, perceptual, acoustic, movement, and neuromuscular. For example, it is not unusual that an utterance is judged "fluent" at the perceptual level by an experienced listener, while analysis at deeper levels of speech, kinematic for instance, indicate inappropriate or "dysfluent" production. Until the distinction between fluent, disfluent, and dysfluent speech is better understood, the adult stutterer's judgment in the classification of him- or herself as a stutterer and in the fluency-dysfluency distinction of his or her speech should be encouraged.

I.B.3. Voluntary and involuntary speech motor output. The distinction between "voluntary" and "involuntary" in defining disfluency and/or dysfluency is not consistently made, although it may represent a critical distinction between certain types of dysfluency exhibited by stutterers and disfluency exhibited by adults who do not stutter. The significance of "involuntary" in definitions of stuttering and stutterers is discussed in detail by Perkins (1983) in response to a review article on stuttering by Andrews and his colleagues (Andrews et al., 1983) and need not be elaborated in great detail here. By way of a brief example, Perkins (1983) writes that the presumption is that a voluntary disfluency in the adult population results from "linguistic uncertainty." This implies that a voluntary disfluency, as in the prolongation of the isolated vowel /a/ for example, is a voluntary strategy invoked by the speaker while attempting to resolve a high-level linguistic query, such as in lexical retrieval. On the hand, the mechanisms underlying involuntary speech acts are far less agreed upon. Citing Perkins (1983) once again as an example, dysfluency "presumably is a motor speech blockage." Certainly, one can argue with linguistic uncertainty and speech motor blockage models of the voluntary-involuntary distinction, but until the volition of speech motor output can be measured with validity, the stutterer's identification of his/her fluent and dysfluent speech should be encouraged. That is, stutterers appear to be in a better position to subjectively rate their utterances as involuntary dysfluent, voluntary disfluent, or fluent than are listener-judges.

These two notions, the stutterer's self-identification and the voluntary-involuntary distinction, are important because they appear to be essential parameters in distinguishing among disfluent, dysfluent, and perceptually fluent speech. The significance of developing criteria to explicitly differentiate among involuntary dysfluency, voluntary disfluency, and perceptual fluency will be discussed in greater detail in Section II.B.3.

I.B.4. Core versus secondary stuttering behaviors. Since many developmental models of stuttering consider some form of repetitions to represent the "core" of stuttering (e.g., Bloodstein, 1987; Stromsta, 1986; Van Riper, 1982) the inclusion of core behaviors in a definition appears warranted.[1] Beyond "involuntary repetitions," and perhaps somewhat secondarily the duration and frequency of prolongations and silent pauses, there appears to be less agreement as to what is sufficient to delimit stuttering, except perhaps for the frequent remark that defining stuttering is much more complicated than some would think. Of course, there is a wide variety of so-called secondary behaviors or accessory features associated with stuttering (e.g., Wingate, 1964), the specification of which would be important especially in the consideration of severity.

I.B.5. Inclusive definitions of stuttering. Although there currently seems to be a better appreciation for psychological effects on physiological behaviors (e.g., Smith & Weber, 1988; Zimmerman, 1980c) definitions for the most part are rarely inclusive of external and internal influences on the disorder. Rather, they are either predominantly psychologically or physiologically based. Comprehensive definitions of stuttering similar to the "integrated theory" notion proposed most recently by Smith and Weber (1988) need to be better developed.

Definitions of stuttering do not always include an account of the suspected etiology of the disorder. For example, definitions could include the notion that involuntary "core" behaviors occur as a consequence of deficits, at various levels, in temporal programming (e.g., Caruso, Abbs, & Gracco, 1988; Kent, 1984), spatial or movement programming (e.g., Zimmerman, 1980c), or both temporal and spatial programming (e.g., Alfonso et al. (1986a; 1987a,b,c).

With a proper definition of stuttering, the researcher is better able to define stuttering subjects. Section II.A. shows, however, that researchers generally provide little definition of the stuttering populations that serve in their experiments, one consequence of which is that it often makes it difficult to make appropriate comparisons among experiments.

SECTION II: SUBJECT DEFINITION AND SELECTION CRITERIA

II.A. An Overview

A review of the literature indicates that little detail is given in journal articles regarding either subject definition or selection criteria. It is more often the case that stuttering severity is reported, although the means by which the severity estimate is determined is highly variable across experiments. What follows are examples of subject selection criteria (and subject definition, if given) that have been reported in recent or frequently cited research papers. The aim is to demonstrate the variability among published descriptions of experimental subjects and the criteria employed to select and define them. The following citations are in alphabetical order.

Many reports of experiments in the contemporary literature provide little information at all. For example, Freeman and Ushijima (1978) state only that their subjects were mild-moderate or severe. No other details of subject selection or definition criteria are given. Guitar et al. (1988) reports the gender of the subjects, that they were all native speakers of the same language, and had never received treatment for stuttering. The first author subjectively judged the severity of the subjects. The criteria for estimating severity were not given. Although the motivation for the Martin and Haroldson (1988) study was to experimentally increase stuttering frequency, so that the definition of stuttering, the frequency of stuttering, and the severity criteria, are crucial in this type of study, very little detail is given. The procedure employed is difficult to ascertain. "In the control room, the experimenter monitored all sessions auditorily and depressed a handswitch each time the subject stuttered. Stuttering was defined in terms of *moments* or *instances* of stuttering and not in terms of specific disfluency types." McClean (1987) reports that "informal assessment of conversations with the stutterers (7 adult male) suggested that as a group their stuttering severity ranged from mild to severe." No other details of subject selection or definition criteria are given. Zimmerman (1980a,b) and Zimmerman and Hanley (1983) used subjects who were enrolled in speech and hearing clinics at the time of data collection. The specific type of therapy is not mentioned. No details of the severity

criteria are given: "They ranged in severity from mild to severe as judged by a certified speech-language pathologist."

Other reports of experiments provide some detail regarding subject selection criteria, though the type and amount vary in considerable degree. For example, Alfonso and his colleagues (Alfonso et al., 1986a,b; 1987a,b,c; Kalinowski & Alfonso, 1987; Story & Alfonso, 1988; Watson & Alfonso, 1982, 1983, 1987) used subjective and objective criteria to identify stutterers and group them on the basis of stuttering severity. Objective evaluations of stuttering frequency and type were completed using a combination of procedures described in the Stuttering Interview (SI) (Ryan, 1974) and the Stuttering Severity Instrument (SSI) (Riley, 1972). Subjective judgments of stuttering severity were obtained from certified speech-language pathologists and from the experimental subjects. Additional criteria are developed if severity ratings differ markedly between reading and conversational speech samples, among objective and subjective criteria (e.g., Watson & Alfonso, 1987), or if inter-test reliability, as a function of time periods or the identity of the judges, was low (Kalinowski & Alfonso, 1987). Caruso et al., (1988) assessed stutterers' dysfluency during a conversational speech sample using two objective measures: 1), mean stuttering frequency (MSF), and 2), mean stuttering duration (MSD). Stuttering was defined as *sound/syllable* repetitions and *sound prolongations*. Conture et al. (1977) used MSF during conversation and oral reading. Stuttering severity was determined by use of the Iowa scale (Johnson, Darley, & Spriestersbach, 1963) based on the MSF. All of the subjects for Conture et al., (1985) were receiving therapy at a university speech and hearing clinic. MSF for sound/syllable repetition, sound prolongation, or within-word pause was used as a measure of severity. Metz et al. (1983) and Sacco et al. (1987) selected subjects enrolled in a residential stuttering treatment program. They used MSF based on sound and syllable repetitions, sound prolongations, and/or broken words produced during an unidentified reading sample. Severity was calculated using the SSI (Riley, 1972; 1980). The Shapiro (1980) experiment required that the locus of stuttering, type of dysfluency, and stuttering severity be reliably determined. Subjects read the Rainbow Passage five times and performed the Job and TAT Tasks (Johnson et al., 1963). Subjects were accepted if the interjudge agreement across four judges meet criteria. Operational definitions of severity of stuttering were obtained using the Stuttering Severity Scale (Johnson et al., 1963), and estimates of the specific type and locus of stuttering were subjectively made from videotape viewing.

II.B. Needs and future directions

The section above indicates that it is usually the case that one would find little detail in a journal article regarding the subject definition and selection criteria employed in the experiment. Stuttering severity estimates are given more often than subject definition and selection criteria, although the means by which the severity estimate is arrived at is highly variable across experiments. In what follows, certain issues relevant to the development of subject definition and selection criteria are discussed in detail. The issues represent broad areas of concern, and should not be construed as suggested minimal criteria. Rather, criteria should be developed as a function of the nature of the experiment at hand. The goal that all researchers should share, however, is that adequate descriptions should be given in sufficient detail so that experiments can be replicated and/or results can be appropriately interpreted and compared among experiments.

II.B.1. What identification measures should be used? The common use of a standardized perceptual test such as the SSI would lead to an obvious advantage of direct comparison of subjects across different experiments. Other identification measures should be considered, and include: (a) familial history of stuttering,[2] (b) type(s) and duration(s) of therapies received, the distinctive characteristics of a therapy program, a description of clinical goals (e.g., slowed speech rate, gentle onsets) of these therapies, and how well they are maintained in habitual speech, (c) estimates of covert stuttering severity, for example, the stutterer's judgment of the frequency and severity of dysfluency, secondary characteristics (e.g., Riley, 1972; Van Riper, 1982), descriptions of contextual conditions in which fluency and dysfluency are enhanced, and an estimate of the success to which the subject is able to use fluency enhancers to promote fluent speech, (d) estimates of overt stuttering severity, for example, the frequency and duration of repetitions, prolongations, and silent pauses, and (e) a description of the subject's fluent speech, for example, rate and naturalness. Speech samples gathered as part of the evaluation should be obtained during both extemporaneous and read speech, and the differentiating stuttering behaviors should be noted. Speech samples

gathered as part of the evaluation and experimental run should routinely be video-taped.

However, perceptual testing alone poses a number of problems. The first is that the magnitude of stuttering severity is highly variable within subjects so that a classification of "severe" could be given to a subject tested in the morning while a classification of "moderate" could be given to the same subject tested in the afternoon (This problem is discussed in greater detail in Section II.B.4.) A second problem, if one accepts the notion of physiological loci, is that available standardized tests do not adequately differentiate between, for example, labial stutterers and laryngeal stutterers. The latter problem is related to a more serious shortcoming of subject selection and classification schemes made on the basis of standardized perceptual testing alone, in that no physiological measurement of speech is attempted. The omission of physiological description is a serious omission because: (a) a physiologically based classification may be more stable across time than a perceptual method, especially as the linguistic structure of dysfluent speech is better understood and the corresponding linguistic-physiological data base is increased, and (b) a dysfluent utterance may not be identified at the perceptual level. That is, while a segment of a stutterer's speech may be judged "fluent" by a group of listeners, the acoustic, and/or movement, and/or electromyographic signals underlying the perceptual segment may appear inappropriate or "dysfluent" (e.g., Alfonso et al., 1984; Baer & Alfonso, 1984; Shapiro, 1980).

A relatively noninvasive physiological-based battery could be devised. The following are examples of possible criteria that could be included in a subject selection and definition battery. Of course, the selection and applicability of physiological criteria are dependent upon the nature of the proposed experiment, the point, however, is that frequent and systematic use of certain physiological measures would enhance the validity of data comparison across different experiments. Noninvasive respiratory criteria, based on data obtained by a Respitrace inductive plethysmograph for example, could include the magnitude of the lung volume exchange for speech inspiration and expiration, or flow rates during fluent speech, since both have been shown to differentiate stutterers from nonstuttering adults, and, perhaps distinguish mild from severe stutterers (e.g., Lewis, 1975; Peters & Boves, 1987, 1988; Watson & Alfonso, 1987). An extensive body of literature on the differences between stutterers' and control subjects' acoustic reaction-time responses (e.g., Watson & Alfonso, 1983, 1987) suggest that predominantly laryngeal (and secondarily respiratory and supralaryngeal) noninvasive criteria could include acoustic reaction-time latencies for steady-state vowel responses. More direct noninvasive laryngeal estimates could be made using an electroglottograph with the reaction-time paradigm. Vocal-tract criteria based on direct measures of supralaryngeal articulator movements would require more technologically sophisticated noninvasive paradigms than those discussed above, and could include, for example, strain-gauge or opto-electronic movement transducers to measure lip and jaw displacement. These techniques are becoming more common in many laboratories. It would be important that supralaryngeal physiological criteria for the selection and definition of stuttering subjects be centered on organizational principles of speech motor control, that is, centered on good representatives of what are believed to be speech motor control parameters. For example, noninvasive lip and jaw displacement amplitude data could be analyzed to assess motor equivalence covariation and sequential ordering in labial gestures (Alfonso et al., 1986a, 1987b,c; Caruso et al., 1988). This issue is discussed in greater detail in Section II.B.3. It is less important to know and base group comparisons on displacement amplitudes and velocities of individual lip and jaw movements, and at what rates they move, than it is to know about the organizational principles underlying lip-jaw movements because (a) efficient, rapid, and fluent speech requires a relatively high degree of spatial and temporal coordination among the supralaryngeal speech structures, and (b) interspeaker variability at the motor control level is inherently less variable than at the phonetic level. Of course, data obtained by a combination of these techniques could be used to measure intersystem physiological parameters, for example, respiratory-laryngeal timing (e.g., Peters & Boves, 1988; Watson & Alfonso, 1987).

II.B.2. What behaviors during the moment of stuttering should be included in the definition? The type, duration, and severity[3] of the involuntary dysfluency should be identified during the moment of stuttering. Because the relationship between linguistic structure and speech motor specification in dysfluent speech is not fully understood, a description of the linguistic context in which the dysfluency occurred should be given. At the very least, the intended fluent phonetic target should be identified, for example, stressed syllable initial voiceless aspirated stop. Other

linguistic descriptors that could be specified depending on the focus of the experiment include word position and content versus function word. A detailed consideration of psycholinguistic variables associated with stuttering is given in Wingate (1988). A number of secondary characteristics, including anxiety and emotional stress, should also be considered.

Although Shapiro (1980) concluded on analysis of EMG data that the perceptions of the judges as well as the subjects themselves regarding the identification of labial, laryngeal, and lingual predominate locus were erroneous, it might still be clinically and theoretically useful, particularly in regard to the subject's perception of his/her production of the dysfluency, to include in the definition a statement about clinician and/or client estimate of locus. We need more physiological data, in parallel with perceptual and acoustic, to determine with certainty whether or not stuttering results from a physiological disruption at one location, the larynx for example, or if the entire speech motor system fails simultaneously. Shapiro concludes that the stuttering may not be able to identify the location of the disruption, however, we don't have enough data to know this for certain, nor do we have enough data to know whether a breakdown at the larynx, for example, occurs first and leads to failures in other components of the speech system in response to the laryngeal failure.

Finally, for those experiments based on physiological data it may be useful to segment the events surrounding a dysfluency. By way of example, Alfonso and Seider (1986) examined acoustic, respiratory and laryngeal kinematic data, and laryngeal electromyographic (EMG) data during an interval beginning with the termination of fluency, followed by an inaudible dysfluent period, followed by an audible dysfluent period, and ending with a fluent period. Laryngeal EMG and movement data showed that the laryngeal configuration was clearly different during the inaudible than during the audible periods of the dysfluent episode. The configuration appeared most inappropriate during the initial inaudible period and less so during the following audible period. Segmentation also allows for the comparison of speech motor events during the dysfluent periods with events immediately following, that is, during fluent production of the phonetic target. Thus, statements about the configuration of the vocal tract during moments of stuttering can be made in reference to the vocal tract configuration during more fluent episodes immediately preceding the dysfluency, and in reference to the vocal tract configuration immediately following the dysfluency and associated with the intended fluent phonetic target.

II.B.3. Is there such a thing as "normal disfluency" and can it be reliably differentiated from stuttering? Perceptual fluency spans a wide continuum, the endpoints of which could be identified as "severely fluent" to "severely dysfluent." There is large variability within the "normal" subsection of the continuum, ranging from something like "severely fluent" to "normal disfluent" and apparently even more variability within the "abnormal" subsection of the continuum, ranging from something like "abnormal fluent" to "severely dysfluent." Because the "normal" and "abnormal" subsections of the continuum appear to overlap perceptually, it seems that defining *abnormal fluency*, particularly when produced by severe stutterers, may be more straightforward than defining *normal disfluency*. That is, it may be more experimentally viable and more fruitful in the long run to modify the above question and to ask: is the perceptual "fluent" speech of severe stutterers similar to the fluent speech of control subjects? The latter form of the question may be more experimentally viable because: (a) the resolution involves extreme contrasts, the continuum endpoints, represented by a severe stutterer contrasted with a fluent control subject in the example here. The extreme contrast should be easier to differentiate than more subtle contrasts; for example, certain physiological characteristics of the perceptually fluent speech of a severe stutterer compared with the corresponding characteristics of a normal subject's fluent speech should be easier to differentiate than a mild stutterer's fluent speech compared with a normal subject's fluent speech, and (b) it would be easier to define a severely dysfluent stutterer than it would be to define a severely fluent control subject. The latter form of the question may be more fruitful because a better understanding of stutterers' perceptually fluent speech would represent a relatively direct and immediate increase in our understanding of stutterers' speech motor control. Of course, in the long run we will need to understand better the fluency variability in the normal population as well as the stuttering population. Thus, the following discussion is a modification of the originally assigned question and asks: (a) can normal disfluency be reliably differentiated from stutterers' dysfluency, and (b) can stutterers'

perceptually fluent speech be reliably differentiated from normal fluent speech?

Considering normal disfluency first, certainly adults who do not stutter do repeat and prolong a variety of speech segments, usually words and phrases, and interject pauses between words and phrases. Generally, the speech segments in which repetitions occur differentiate the groups; nonstutterers predominantly repeat whole words and phrases whereas stutterers predominantly repeat sounds and syllables. The duration of pauses and prolongations may also distinguish normal disfluency from abnormal dysfluency. However, the frequency of repetitions and the duration of prolongations, regardless of the speech segment in which these occur, are far less in magnitude in the nonstuttering population than in the stuttering population. That is, stutterers do more of everything; repetitions, prolongations, and pauses. Normal disfluencies are not usually accompanied by secondary characteristics and may be less susceptible to higher linguistic influences, for example, word order and word type (see Starkweather, 1987, Chapter 5 for a more detailed discussion of normal disfluency). An important assumption underlying the distinction between normal and abnormal speech is that normal speech movements are voluntary whereas stutterers' dysfluencies are not. Although it may be difficult, and perhaps impossible, to ascertain the volition of speech motor output, determining the extent to which repetitions, prolongations, and pauses are under voluntary motor control may be the ultimate test of the normal disfluency versus abnormal dysfluency distinction.

Considering stutterers' perceptually fluent speech next, there is considerably more data regarding the contrast between stutterers' perceptually fluent speech and normal fluent speech than in the contrast discussed above. However, some experimenters conclude that stutterers' perceptually fluent speech is not different than control subjects' fluent speech, while the majority of experimenters seem to think that it is (see, for example, Van Riper, 1982, Chapter 16; Bloodstein, 1987, Chapter 1). Rather than review a relatively large literature here, it may be more useful in regards to the development of subject definition and selection criteria to discuss a few of the reasons underlying the conflicts in the results of these types of experiments. It should be noted first, however, that the conflicting results are no doubt confounded by the lack of adequate definitions and criteria for distinguishing among certain essential characteristics of stuttering discussed in Section I, for example, voluntary disfluency, involuntary dysfluency, and perceptual fluency. That is, the results of two experiments may generate conflicting conclusions whether or not stutterers' perceptually fluent speech is similar to normally fluent speech simply because the criteria (which may or may not be reported) for distinguishing the stutterers' perceptually fluent speech from their dysfluent speech differed across the two experiments.

Second, the majority of experiments are based on perceptual data alone, some are based on perceptual and speech acoustic data, and far less include kinematic and neuromuscular data in parallel with perceptual and acoustic. One source of the conflict was discussed above in Section II.B.1, that is, while a segment of a stutterer's speech may appear normal or "fluent" at the perceptual level, it may appear abnormal or "dysfluent" at the acoustic, and/or movement and muscular levels (Alfonso et al., 1984; Baer & Alfonso, 1984; Shapiro, 1980). Thus, a comparison based on perceptual data alone could indicate no difference between the groups, while a comparison of the same utterances using the same perceptual criteria for fluency and dysfluency but based on physiological data could find significant group differences. Clearly, perceptual data alone are too far removed from the source to be able to make detailed analyses of the fluent-dysfluent distinction, and as such would mark only the most obvious instances of stuttering. The lack of physiological data addressing this issue is of concern for other reasons. For example, the question of determining whether the perceptually "fluent" speech of severe stutterers is similar to the fluent speech of control subjects is important because the answer to the question will help determine whether the stutterers' speech motor system exhibits generalized spatio-temporal abnormalities regardless of the perceived fluency, or whether, alternatively, it behaves normally except during moments of dysfluency. It is generally the case that perceptual and acoustic data in the absence of simultaneously gathered physiological data are not sufficient to answer questions about speech motor control.

A second reason for the conflict in the results of reported literature on the distinction between stutterers' perceptually fluent speech and nonstutterers fluent speech is that the routinely posed form of the question, "Is stutterers' fluent speech similar to the fluent speech of adults who do not stutter," is too broad. It is possible that

certain aspects of speech do not differentiate the groups while other aspects do differentiate the groups. For example, Baken et al. (1983) found no differences between stutterers and control subjects in chest wall preposturing maneuvers immediately preceding fluent speech. Yet, significant differences in subglottic pressure (e.g., Lewis, 1975), flow rates (e.g., Hutchinson, 1975), and lung volume charge and deflation for speech (e.g., Story, 1990; Watson & Alfonso, 1987) have been observed between the groups. Thus, it is less appropriate to ask whether or not respiratory control, per se, (and certainly not speech production, in general), can be differentiated between the groups. Rather, it might be that stutterers and normal speakers perform certain speech respiratory gestures in a similar fashion, namely respiratory preposturing, but that other aspects of speech respiration, namely the magnitude of the inspiratory charge, are performed differentially.

A third and perhaps most important reason for the conflict in the results of these types of experiments is that group comparisons are frequently based on physiological data that reflect relatively variant phonetic level speech gestures. Group data based on phonetic level contrasts are inherently unstable since spatial and temporal control of the speech structures to mark phonetic distinctions varies as a function of phonetic context, stress and rate, dialect, and individual speaker preferences. Rather, group comparisons should be based on relatively stable spatial and temporal characteristics of normal speech dynamics, for example those that best reflect organization principles of speech motor control. Dynamic parameters that are relatively invariant across multiple productions of an utterance are thought to be good representatives of speech motor control parameters (e.g., Gracco & Abbs, 1986). Thus, an appropriate comparison would be one based on the extent to which the magnitude of the relatively stable spatial and temporal characteristics of normal speech dynamics approximates corresponding characteristics of stutterers' perceptually fluent speech (Alfonso et al., 1986a, 1987b,c; Caruso et al., 1988). For example, motor equivalence covariation for speech is based on the observation that normal subjects control the relative displacement of individual articulators (e.g., the tongue and jaw) enlisted in a vocal tract gesture (e.g., alveolar closure) in such a way that the variability of the gesture is less than the variability associated with the individual articulators comprising the gesture. Accordingly, it is less appropriate to base group comparison on displacement amplitude and peak velocity of a single articulator, the tongue or the jaw for example, than it would be to compare the relative organization of the combined tongue-jaw displacement, because the latter comparison reflects a relatively stable component of speech dynamics while the former reflects idiosyncratic speaker preference. To summarize, the question of whether stutterers' perceptually fluent speech is similar to the fluent speech of non-stuttering adults can be appropriately addressed if criteria are developed that: (a) objectively distinguish among perceptually fluent, disfluent, and dysfluent utterances, (b) define the question more precisely, that is, provide more detail about specific aspects of speech production, (c) require appropriate types of data, and (d) require group comparisons based on appropriate speech behaviors.

II.B.4. How can a definition take into account variability in stuttering behavior? At least three types of variability should be taken into account. The first is associated with the well-known variability in the frequency and severity of dysfluent behaviors as a function of various external conditions. For example, dysfluent behaviors decrease and perceptually fluent behaviors increase with manipulation of auditory feedback; when the intensity of the feedback signal is increased, decreased, or masked by various techniques, and when the time of arrival of the auditory feedback signal to the talkers' ears is delayed. Fluency is enhanced by imposing an external rhythmical marker on the speech task, for example, a metronome signal, choral reading, and singing. Finally, fluency is enhanced by the adaptation effect. Most of these conditions will enhance fluency in non-stuttering adults as well as stutterers. The exception seems to be the delayed auditory feedback (DAF) paradigm, where a delay in the signal could result in increased dysfluency in some non-stuttering adults, although normal speakers vary widely in their susceptibility to DAF (Alfonso, 1974). Thus, alterations of the feedback signal, imposition of a rhythmical marker, and multiple repetitions of the same speech task increase fluency in most talkers, stutterers and nonstutterers alike. It appears possible that all of these conditions enhance fluency through a similar mechanism, that is, they act to highlight the control of prosody and perhaps other temporal parameters of the speech motor plan.

Likewise, there are a number of external conditions that increase dysfluent behaviors and

decrease perceptually fluent behaviors. It is well known that the contextual surround, for example, stressful situations and tasks, and specific listeners, will increase the frequency and severity of dysfluency. Certain linguistic conditions, such as long words, content versus function words, and words carrying high information loading, are known to promote dysfluency. Once again, these conditions are known to similarly affect adult nonstutterers, though not to the degree seen in stutterers. On the other hand, there are those adult nonstutterers (and apparently stutterers) who become more fluent in seemingly stressful situations. A likely causal agent for the increase in dysfluency in these conditions for both stutterers and adults who do not stutter is the effect of stress on the speech motor system (e.g., Zimmerman, 1980c).

The second type of variability is associated with changes in the type, frequency, and/or severity of dysfluent behaviors that occur across varying units of time. As an example of the variation in dysfluent behaviors that occur across the long term, it is generally observed that very young children who stutter do so primarily by repetition and prolongation, while over the course of 2 to 4 years develop a variety of other overt behaviors not present at the onset. Thus, stuttering in adults is comparatively idiosyncratic and as such varies considerably across the adult stuttering population. There are other examples of long-term changes in dysfluent behaviors, although the basis for the change is not sufficiently understood. With respect to the example given here, the source and extent of the variability of "core" behaviors, the repetitions and prolongations in very young children, compared to the variability of "secondary" behaviors in older children and adults apparently remains unresolved (Stromsta, 1986; Van Riper, 1982). Of course, variations in dysfluent behaviors occur across much smaller units of time. In fact, the perception of stuttering frequency, severity, and type within the same subject can change significantly as a function of many of the external conditions mentioned above (e.g., stressful situations and tasks) over the course of hours and days. Thus, as was pointed out in Section II.B.1., the same subject can be classified as a mild-moderate stutterer in a morning session and using the same severity rating instrument be classified as a severe-moderate stutterer in an afternoon session. This poses a particular problem in the development of subject definition and selection criteria and illustrates the importance of accounting for the variety of conditions that are known to induce short- and long-term fluency variations.

A third type of variability that could be included in subject definition and selection criteria is the type observed at the speech motor level. Greater variability for stutterers compared to control subjects at every accessible level of speech production, namely, acoustic, movement, and neuromuscular, is frequently reported and must be considered a robust observation. Relatively high levels of variability can, of course, be of serious consequence in meeting the demands of rapid, fluent speech. High variability implies that stutterers' control of the movements of the speech structures is less precise than adults who do not stutterer. The lack of precision implies that stutterers lack the flexibility observed in normal speech, the flexibility that forms the basis for co- production of speech segments and in other human multiarticulate systems. Thus, the increased variability observed in the kinematics of many structures of the speech mechanism, and more importantly, the increased variability associated with various organizational components of fluent speech (e.g., the tongue-jaw synergies referred to in Section II.B.3.) suggest that stutterers' control of the speech motor system lacks the flexibility and efficiency to meet the demands of rapid, fluent speech. (e.g., Alfonso et al., 1985, 1987c; Stromsta, 1986).

In summary, subject definition and selection criteria should account for the variability that is frequently observed at all levels of stuttering behaviors. In regards to treatment of the data, descriptions of data dispersion in addition to central tendency should be reported. Reporting the complete data base, perhaps as an appendix, would often be appropriate. Statements regarding the suspected source or the conditions promoting the variability, and the effect of the source or conditions on the variability, should be included; for example, fluency enhancement by rhythmical stimulation, dysfluency increase by psychological and physiological stress, accompanied by a statement regarding the magnitude of the fluency change. Finally, we need to understand better the relationship between the magnitude of the variability and stuttering severity.

II.B.5. What are the commonalities among stutterers? There are certain characteristics that appear to be common among stutterers as a group and should be considered in developing subject definition and selection criteria. For example, all stutterers seem to experience the notion of stutterer-identification discussed above in Section I.B.2. That is, regardless of severity, or the degree

to which overt or covert manifestations of stuttering are realized, stutterers seem to know that they are stutterers. If this is indeed true, then the notion of self-identification should be a central component of the definition of stuttering and stutterers. The stutterers' identification of themselves as stutterers and their identification of their moments of stuttering should focus the development of perceptual and physiological criteria.

Based on family history data, many researchers have concluded that the predisposition to stutter is genetically transmitted. For example, Van Riper (1982) concludes: "This incidence is so much greater than that found generally in the population (about 5 per cent) that it is difficult to believe that environmental factors alone could account for the results" (p. 330). Family history studies show that more males than females stutter, that there is a high incidence of stuttering within families, and that the risk of stuttering is higher if the mother, rather than the father, stutters. Certain of these studies allow the determination of risk to relatives; stuttering occurred in about 20% of male relatives and about 5% of female relatives of male stutterers, whereas stuttering occurred in about 25% of male relatives and 12% of female relatives of female stutterers (Kidd et al., 1978, 1981). However, as Pauls (1990) points out in a paper entitled "A review of the evidence for genetic factors in s11

tuttering," there are significant limitations to the family study method. Coupled with the fact that these limitations make it impossible to understand the exact nature of the transmission, and that a disorder as common as stuttering is almost certain to be etiologically and genetically heterogeneous, the inclusion of family history data in subject definition and selection criteria should be cautiously considered.

The onset of stuttering is fairly common across stutterers. Most stuttering develops during childhood and is usually marked initially by syllable repetitions and prolongations. Adult stutterers who demonstrate a characteristic childhood onset should be distinguished from the relatively few adult stutterers who report sudden onset later in life.

Finally, certain physiological characteristics of the speech motor system are shared by most stutterers. For example, the stutterers' speech motor system is inappropriately susceptible to psychological and physiological stress. Fluency, measured at the output of the system primarily by perceptual criteria, is enhanced by rhythmical stimulation. The reaction-time of the stutterers' speech motor system is slower than normal speakers. And the stutterers' system may be characterized by a disability in spatial, temporal, or spatio-temporal coordination, particularly of the type discussed in Section II.B.3.

II.B.6. What would be the benefits to research in having a consensus definition? While it is likely that reaching the ultimate consensus definition is not possible to achieve, the alternative, that is to continue without suggestive guidelines would lead to a continuation of the variability in subject definition and selection criteria illustrated in Section II.A. One of the aims of this paper is to demonstrate the different ways in which current definitions and c[illegible]eria are inadequate. With respect to definitio[illegible] of stuttering, there are many instances where data have been pooled across such factors as severity and type so that a clear interpretation of the results is not possible. The lack of clearly stated subject selection criteria may have even more serious consequence on data interpretation because of the heterogeneous nature of the disorder. For example, data are frequently pooled across such variables as subject severity, varying clinical treatment influences on perceptually fluent speech patterns, and idiosyncratic dysfluent verbal and nonverbal behaviors.

There are, however, obvious and significant limitations to the consensus approach. The first is related to the wide varieties of stuttering behaviors and subjects that are often examined. Overly precise definitions may exclude critical members of a category while overly general definitions would not make the necessary distinctions between members of a category. Secondly, standardized subject definition and selection criteria would be in a constant state of reassessment as our knowledge of the disorder continues to increase.

Many of the benefits of a consensus definition can be achieved by the routine practice of reporting sufficient details of the experiment so that, at the very least, the experiment can adequately be replicated. The precise details will vary as a function of the nature of the experiment, yet it would be difficult to imagine many cases where basic descriptors, for example, severity ratings of the dysfluencies and subjects under examination, could be omitted. The paper presented here discusses a number of factors that bear on subject definition and selection criteria. Undoubtedly, other factors not listed here could also be included. Thus, the consensus opinion should be that subject definition and selection

criteria adequately deal with the heterogeneity of stuttering, that they be specifically generated as a function of the experiment at hand, and that until the significance of the heterogeneity is better understood, definitions and criteria should err on the side of over- rather than under-defining critical details of the experiment. Included in the notion of over-defining in the context of a heterogeneous behavior is the treatment of the data. Because the reported data are often used by future researchers in a way not related to the aim of the original experiment, it would be appropriate to publish certain forms of individual subject data as an appendix, in addition to the typical reporting of averaged data, especially when the data represent physiological estimates of stuttering that are technically difficult to collect and/or potentially hazardous to the subject. The consensus, then, becomes one of intent rather than the specification of consensus criteria.

REFERENCES

Alfonso, P. J. (1974). *Differentiating phonemic and spectrographic speech characteristics of DAF susceptibility*. Unpublished M.A. thesis, Western Michigan University.

Alfonso, P. J., Watson, B. C., & Baer, T. (1984). Muscle, movement, and acoustic measurements of stutterers' laryngeal reaction times. In M. Edwards (Ed.), *Proceedings of the 19th Congress of the International Association of Logopedics and Phoniatrics* (Vol. II, pp. 580-585). Perth, Scotland: Danscott Print Limited.

Alfonso, P. J., Watson, B. C., Baer, T., Sawashima, M., Hirose, H., Kiritani, S., Niimi, S., Itoh, K., Sekimoto, S., Honda, K., & Imagawa, H. (1985). The organization of supralaryngeal articulation in stutterers' fluent speech production: A preliminary report. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics, 19*, 191-200.

Alfonso, P. J., Watson, B. C., Seider, R. A. (1986a). Tongue and jaw articulation in stutterers' fluent speech. *ASHA, 28-10*, 166.

Alfonso, P. J., & Seider, R. A. (1986b). Laryngeal and respiratory physiological characteristics of inaudible and audible dysfluent production. In S. Hibi, D. Bless, & M. Hirano (Eds.), *Proceedings of the International Congress on Voice* (1, pp. 13-22). Kurume, Japan: Kurume University.

Alfonso, P. J., Watson, B. C., & Baer, T. (1987a). Measuring stutterers' dynamical vocal-tract characteristics by x-ray microbeam pellet-tracking. In H. F. M. Peters & W. Hulstijn (Eds.), *Speech motor dynamics in stuttering* (pp. 141-150). Wien/New York: Springer-Verlag Press. .

Alfonso, P. J., Story, R. S., & Watson, B. C. (1987b). The organization of supralaryngeal articulation in stutterers' fluent speech production: A second report. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics, 21*, 117-129.

Alfonso, P. J., Story, R. S., & Watson, B. C. (1987c). Spatial-temporal variability of tongue- jaw movements in stutterers' fluent speech. *ASHA, 29-10*, 159.

Andrews, G., Craig, A., Feyer, A. M., Hoddinott, S., Howie, P., & Neilson, M. (1983). Stuttering: A review of research findings and theories circa 1982. *Journal of Speech and Hearing Disorders, 48*, 226- 246.

Baer, T., & Alfonso, P. J. (1984). On simultaneous neuromuscular, movement, and acoustic measures of speech articulation. In R. G. Daniloff (Ed.), *Articulation assessment and treatment issues*. San Diego: College Hill Press.

Baken, R. J., McManus, D. A., & Cavallo, S. A. (1983). Prephonatory chest wall posturing in stutterers. *Journal of Speech and Hearing Research, 26*, 444-450.

Bloodstein, O. (1987). *A handbook on stuttering*. Chicago, IL: The National Easter Seal Society.

Caruso, A. J., Abbs, J. H., & Gracco, V. L. (1988). Kinematic analysis of multiple movement coordination during speech in stutterers. *Brain, 111*, 439-455.

Conture, E. (1990). Childhood stuttering: What is it and who does it? *ASHA, 18*, 2-14.

Conture, E., McCall, G., & Brewer, D. (1977). Laryngeal behavior during stuttering. *Journal of Speech and Hearing Research, 20*, 661-668.

Conture, E., Schwartz, H. D., & Brewer, D. (1985). Laryngeal behavior during stuttering: A further study. *Journal of Speech and Hearing Research, 28*, 233- 240.

Fairbanks, G. (1960). *Voice and articulation drillbook*. New York: Harper & Row.

Freeman, F. J., & Ushijima, T. (1978). Laryngeal muscle activity during stuttering. *Journal of Speech and Hearing Research, 21*, 538-562.

Gracco, V. L., & Abbs, J. H. (1986). Variant and invariant characteristics of speech movements. *Experimental Brain Research, 65*, 156-166.

Guitar, B., Guitar, C., Neilson, P., O'Dwyer, N., & Andrews, G. (1988). Onset sequencing of selected lip muscles in stutterers and nonstutterers. *Journal of Speech and Hearing Research, 31*, 28-35.

Ham, R. E. (1989). What are we measuring? *Journal of Fluency Disorders, 14*, 231-243.

Hutchinson, J. (1975). Aerodynamic patterns of stuttered speech. In L. M. Webster & L. C. First (Eds.), *Vocal tract dynamics and dysfluency*. New York: Speech and Hearing Institute.

Johnson, W., Darley, F., & Spriestersbach, D. (1963). *Diagnostic methods in speech pathology*. New York: Harper and Row.

Kalinowski, J. S., & Alfonso, P. J. (1987). Improvement in laryngeal reaction time with improvement in fluency. *ASHA, 29-10*, 124.

Kent, R. D. (1984). Stuttering as a deficiency in temporal programming. In W. H. Perkins & R. Curlee (Eds.), *Nature and treatment of stuttering: New directions*. San Diego: College Hill Press.

Kidd, K. K., Kidd, J. R., & Records, M. A. (1978). The possible causes of the sex ratio in stuttering and its implications. *Journal of Fluency Disorders, 3*, 13-24.

Kidd, K. K., Heimbuch, R. C., & Records, M. A. (1981). Vertical transmission of susceptibility to stuttering with sex-modified threshold. *Proceedings of the National Academy of Sciences, 78*, 606-610.

Lewis, J. J. (1975). *An aerodynamic study of 'artificial' fluency in stutterers*. Unpublished doctoral dissertation, Purdue University.

Ludlow, C. L. (1990). Research procedures for measuring stuttering severity. *ASHA, 18*, 26-31.

Martin, R. R., & Haroldson, S. K. (1988). An experimental increase in stuttering frequency. *Journal of Speech and Hearing Research, 31*, 272-274.

Metz, D. E., Samar, V. J., & Sacco, P. R. (1983). Acoustic analysis of stutterers' fluent speech before and after therapy. *Journal of Speech and Hearing Research, 26*, 531-536.

McClean, M. D. (1987). Surface EMG recording of the perioral reflexes: Preliminary observations on stutterers and nonstutterers. *Journal of Speech and Hearing Research, 30*, 283-287.

Pauls, D. L. (1990). A review of the evidence of genetic factors in stuttering. *ASHA, 18*, 38.

Perkins, W. H. (1983). The problem of definition: Commentary on 'stuttering.' *Journal of Speech and Hearing Disorders, 48*, 246-249.

Perkins, W. H. (1984). Stuttering as a categorical event: Barking up the wrong tree—reply to Wingate. *Journal of Speech and Hearing Disorders, 49*, 431-433.

Peters, H. F. M., & Boves, L. (1987). Aerodynamic functions in fluent speech utterances of stutterers in different speech conditions. In H. F. M. Peters & W. Hulstijn (Eds.), *Speech motor dynamics in stuttering*. Wien: Springer Verlag.

Peters, H. F. M., & Boves, L. (1988). Coordination of aerodynamic and phonatory processes in fluent speech utterances of stutterers. *Journal of Speech and Hearing Research, 31*, 352-361.

Riley, G. (1972). A stuttering severity instrument for children and adults. *Journal of Speech and Hearing Disorders, 37*, 314-322.

Riley, G. (1980). *Stuttering severity instrument*. Tigard, Oregon: CC Publications.

Ryan, B. (1974). *Programmed therapy for stuttering in children and adults*. Springfield, IL: Charles C. Thomas.

Sacco, P. R., & Metz, D. E. (1987). Changes in stutterers' fundamental frequency contours following therapy. *Journal of Fluency Disorders, 12*, 1-8.

Shapiro, A. I. (1980). An electromyographic analysis of the fluent and dysfluent utterances of several types of stutterers. *Journal of Fluency Disorders, 5*, 203-231.

Smith, A., & Weber, C. (1988). The need for an integrated perspective on stuttering. *ASHA, 30*-32.

Starkweather, C. W. (1987). *Fluency and stuttering*. Englewood Cliffs, NJ: Prentice-Hall.

Story, R. S. (1990). *Pre- and post-therapy comparison of respiratory, laryngeal, and supra laryngeal kinematics of stutterers' fluent speech*. Unpublished doctoral dissertation, University of Connecticut, Storrs.

Story, R. S., & Alfonso, P. J. (1988). Changes in respiratory kinematics following an intensive stuttering therapy program. *ASHA, 30*-32.

Stromsta, C. (1986). *Elements of stuttering*. Oshtemo, Michigan: Atsnorts Publishing.

Van Riper, C. (1982). *The nature of stuttering* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Watson, B. C., & Alfonso, P. J. (1982). A comparison of LRT and VOT values between stutterers and nonstutterers. *Journal of Fluency Disorders, 7*, 219-241.

Watson, B. C., & Alfonso, P. J. (1983). Foreperiod and stuttering severity effects on acoustic laryngeal reaction time. *Journal of Fluency Disorders, 8*, 183-205.

Watson, B. C., & Alfonso, P. J. (1987). Physiological bases of acoustic LRT in nonstutterers, mild stutterers, and severe stutterers. *Journal of Speech and Hearing Research, 30*, 434-447.

Wingate, M. E. (1964). A standard definition of stuttering. *Journal of Speech and Hearing Disorders, 29*, 484-489.

Wingate, M. E. (1984a). Definition is the problem. *Journal of Speech and Hearing Disorders, 49*, 429-431.

Wingate, M. E. (1984b). Fluency, disfluency, dysfluency and stuttering. *Journal of Fluency Disorders, 9*, 163-168.

Wingate, M. E. (1988). *The structure of stuttering: A psycholinguistic analysis*. New York: Springer- Verlag.

World Health Organization (1977). *Manual of the international statistical classification of diseases, injuries, and causes of death* (Vol. 1). Geneva: World Health Organization.

Zimmerman, G. (1980a). Articulatory dynamics of fluent utterances of stutterers and nonstutterers. *Journal of Speech and Hearing Research, 23*, 95-107.

Zimmerman, G. (1980b). Articulatory behaviors associated with stuttering: Cinefluorographic analysis. *Journal of Speech and Hearing Research, 23*, 108- 121.

Zimmerman, G. (1980c). Stuttering: A disorder of movement. *Journal of Speech and Hearing Research, 23*, 122- 136.

Zimmerman, G., & Hanley, J.M. (1983). A cinefluorographic investigation of repeated fluent productions of stutterers in an adaptation procedure. *Journal of Speech and Hearing Research, 26*, 35- 42.

FOOTNOTES

*Appears in J. Cooper (Ed.), *ASHA REPORTS* monograph "Research Needs in Stuttering: Roadblocks and Future Directions" (18, pp. 15-24). Rockville, MD: American Speech-Language-Hearing Association (1990).

†Also University of Connecticut, Storrs.

[1]However, a paper entitled "Research procedures for measuring stuttering severity" by Ludlow (1990) takes a different approach to the core versus secondary behavior distinction.

[2]For a discussion of the problems associated with gathering familial history data, see a paper entitled "A review of the evidence of genetic factors in stuttering" by Pauls (1990).

[3]For a detailed discussion of the problems associated with severity estimates see a paper entitled "Procedures for assessing the severity of stuttering for research" by Ludlow (1990).

Haskins Laboratories Status Report on Speech Research
1991, SR-105/106, 243-246

Vocal Fundamental Frequency Variability in Young Children: A Comment on *Developmental Trends in Vocal Fundamental Frequency of Young Children* by M. Robb and J. Saxman*

Margaret Lahey,† Judy Flax,†† Katherine Harris,††† and Arthur Boothroyd††††

One of the major finding of Robb and Saxman (1985) regarding mean fundamental frequency (F_0) of children aged 11 to 25 months was the high variability found among utterances particularly for the youngest children (i.e., the 11-16-month-olds). Robb and Saxman hypothesized that such variability may have been related to the onset of purposeful communication. The data presented below further refine this hypothesis and suggest a strong relationship between the F_0 and the communicative function on infants' vocalizations.

In a study designed to see of prosodic variables were related to communicative functions (Flax, 1986; Flax, Lahey, Harris, & Boothroyd, in press), we videotaped 1-hour mother-child interactions of 3 normally developing children (AL, AB, & RS) at three points in time. At Time 1 the mothers reported that the children were producing noncrying vocalizations that could be interpreted as serving some communicative function; at Time 2, the mothers' diary entries indicated a ten-word vocabulary; at Time 3 the entries showed a vocabulary of about 50 words. The ages sampled ranged from about 11 months to about 14 months (see Table 1).

The video recordings were made using a portable camera (GE, ICVA4035E) and video cassette recorder (GE, ICVD4040X). In order to provide signals of appropriate quality for subsequent pitch extraction, additional audio recordings of the child's utterances were simultaneously made using a Realistic FM wireless microphone (32-122OT) and an audio cassette recorder (Sony, TC-FX22). The transmitter of the wireless microphone was placed in a small pocket attached to the back of a zippered vest worn by the child. The vest was an adaptation of the "telebib" developed by Bauer (Kent & Bauer, 1985). The microphone itself was clipped to the inside front of the vest at the neck. The child's mother also wore a lapel microphone that was connected directly to the second channel of the cassette recorder.

Table 1. *Age in months and days, number of vocalization for each child during each session, and mean fundamental frequency F_0, in Hz.*

	Time 1			Time 2			Time 3		
Child	Age*	# Voc.	Mean F_0	Age*	# Voc.	Mean F_0	Age*	#Voc.	Mean F_0
AL	14,24	180	355 (50)	16,29	161	368 (99)	20,05	179	327 (53)
AB	12,00	149	355 (79)	15,21	145	300 (56)	18,27	187	362 (70)
RS	11,09	169	370 (64)	15,04	169	350 (66)	22,05	194	342 (69)

( ) = Standard deviation. * months, days

The audio recordings of child and mother were analyzed using the F_0 extraction circuit of a Kay Elemetrics Visipitch (Model 6087). To facilitate measurement and storage, the "pitch" output of the Visipitch, which consisted of a time-varying DC signal whose voltage was monotonically related to fundamental frequency, was sampled and digitally stored in an Apple II+ computer. Digital to analog conversion was accomplished with a general purpose interface (Cyborg, ISAAC 90A). To aid in segmentation and interpretation, the amplitude envelope of the signal, extracted by means of a custom-built circuit was similarly sampled and stored. Using a software package developed by one of us (AB) for this application, F_0 (see Figure 1) and amplitude tracings for each utterance were shown graphically on the computer monitor as functions of time. Key points on the F_0 contour were selected with a cursor and marked (onset and offset of intonation contour determined by the amplitude tracing). The printout listed numerical F_0 values for each cursor setting. Among the values extracted were the highest and lowest value of F_0 and the onset F_0. The average of the highest and lowest F_0 was computed for each utterance and referred to as the center F_0.

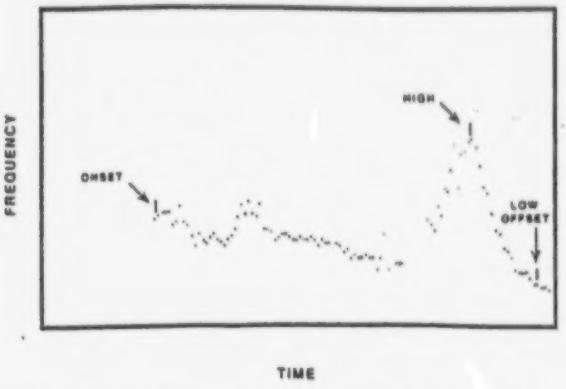


Figure 1. Printout of Visi-pitch tracing of F_0 contour.

Using the video recordings, each vocalization was assigned a contextual function that was an interpretation of how the utterance functioned in the context based on the child's and the mother's behaviors (see Flax, 1986; Flax et al., in press). The function categories used are listed in Table 2.

Table 2. *Contextual function categories.*[1]

Category	Definition-child vocalized:
Response (RESP)	immediately following the question of another
Comment-Label (COM-L)	in apparent recognition of, or to label, a person or object
Request-Attention (RQ-A)	to get attention from another usually a call
Request-Object or Action (RQ-OA)	in accompaniment with reaching, looking, pointing and followed by recognition of request
Request-Response (RQ-R)	and the vocalization was followed by verbal response
Give	as child gave object to another
Request-Command (RQ-C)	in loud voice and repeated until mother complied
Protest (PRO)	in context of unfulfilled desire
Comment-Noninteractive (COM-NI)	in context without gaze or other behavior that indicated it was directed to another
Comment-Interactive (COM-I)	in context where posture and other behaviors indicated an interaction with another person—and vocalization did not fit other categories.

[1]Adapted from Dore (1974), Gerber (1987), and Halliday (1975).

Reliability of contextual function judgments was computed on 20% of the corpus by having an independent coder judge randomly selected segments following a training session. Overall agreement on an item-by-item comparison was 88% with a range for selections from the nine tapes of 83% to 95%. As can be seen in Figure 2, the center F_0 varied with category of contextual function. The center F_0 for categories associated with high emotion such as Protests and Calls for Attention was higher than the center F_0 used for Comments or Responses.

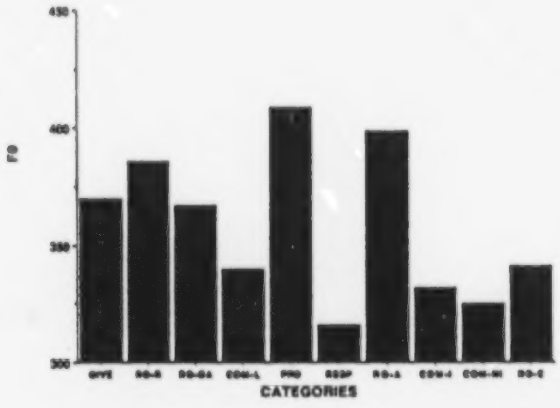


Figure 2. Center F_0 presented by category of contextual function as defined in Table 2.

As with much of child language research, large amounts of data were available here from a small number of subjects and conclusions are limited to the population studied, albeit with implications for further research when consistencies are found among the subjects. With this in mind, child as well as function and time were considered as fixed effects and a three-way analysis of variance (function × time × child) was computed for center F_0. The contextual functions included were those that were used at least five times in each sample (i.e., Comment-Interactive, Request for Object or Action, and Response). A significant main effect was found for function [$F(2,216) = 13.53$, $p < .001$] and a significant interaction was found for Child and Time [$F(4,216) = 7.97$, $p < .001$]. A post hoc analysis using pooled error variance revealed that the effect of time was significant for 2 of the children (AL and AB) but not for the 3rd (RS).

Like Robb and Saxman, we did not find a significant decrease in center F_0 over time. Age was not significantly correlated with center F_0 (rho -.45, $t(7) = 1.35$, $p > .05$). Robb and Saxman (1985) reported a decrease in F_0 when they compared their younger subjects (11-16 months) with their older subjects (17-25 months), suggesting a trend toward lower mean F_0 after the age of 16 months. The longitudinal data on 2 of our 3 children were consistent with this pattern. One child (RS) showed a consistent decrease in center F_0 with age (from 11, to 15, to 22 months) and the other child (AL) evidence a lower center F_0 after 16 months. However, the 3rd child (AB) did not follow this trend and produced a higher center F_0 (362 Hz) at 20 months than at 15 (300 Hz) or 12 (355 Hz) months of age (see Table 1). In a study of 19 infants during their first year of life, Delack (1976) also did not find a decrease in F_0 over time; he reported that mean F_0 was fairly stable over this time period.

Our findings support those of Robb and Saxman (1985) in a number of ways. First, our overall center F_0 of 348 Hz is similar to the mean F_0 reported by Robb and Saxman (357 Hz) for children of similar ages (and to the mean F_0 of 355 Hz reported by Delack, 1976, for 19 children under 1 year). Second, like Robb and Saxman's finding that onset F_0 was similar to mean F_0, we found that onset F_0 (340 Hz) was similar to center F_0 (348 Hz). These similarities in results were found despite differences in methodology. We computed a central F_0 based on the average of the high and low F_0 whereas they sampled 10 data points from each vocalization to obtain a mean; we examined 130 to 180 utterances per sample whereas they examined 70. Furthermore, the instrumental techniques were quite different.

Our results are, however, different from those of Robb and Saxman in one respect. We see no age-related trend in F_0 variability as measured either by the standard deviation or the coefficient of variation. In fact, our results are comparable with those of all but 4 of Robb and Saxman's youngest children (i.e., for subjects 5-14 who were all 15 months or older). This was so even though two of our samples were obtained from children whose ages were similar to Robb and Saxman's youngest children. It is possible that the difference in method and instrumentation is responsible for this difference. More data are needed in the younger age ranges.

Although the variability in our data was not quite the same as theirs, our data do support their hypothesis that variability in F_0 may be related to the onset of meaningful communication. However, in our data it was not the onset of first conven-

tional words that was important. By design, our first samples, at Time 1, were collected before the children were using conventional words; there was little difference in center F_0 between these samples and later samples taken in the single-word-utterance period. Similarly, the center F_0 of vocalizations that were linguistic (i.e., were conventional words) and those that were not linguistic (i.e., not conventional words) was similar in Time 2 (linguistic 343 Hz, nonlinguistic 340 Hz) and Time 3 (linguistic 344 Hz, nonlinguistic 345 Hz). Rather, as shown in Figure 2 it was the association of prosodic variables with various communicative functions that influenced inter-utterance variability of F_0. Thus, what may appear to be minor differences among samples of children's vocalizations might have substantial effects on the distribution of contextual functions expressed by the child, and hence, the observed results. These findings suggest that reports of F_0 and variability in F_0 in this age range and comparison across ages may need to be based on large samples and take into account the types of communicative functions expressed by the children in the samples obtained.

REFERENCES

Delack, J. B. (1976). Aspects of infant speech development in the first year of life. *The Canadian Journal of Linguistics/Le Revue Canadienne de Linguistique, N* 21, 17-37.

Dore, J. (1974). A pragmatic description of early language development. *Journal of Psycholinguistic Research*, 3, 343-350.

Flax, J. (12986). *Functional intonation in the prelinguistic and early linguistic child.* Unpublished doctoral dissertation, The City University of New York.

Flax, J., Lahey, M., Harris, K., & Boothroyd, A. (in press). Relations between prosodic variables and communicative functions. *Journal of Child Language.*

Gerber, S. (1987). *Form and function in early language development.* Unpublished doctoral dissertation, The City University of New York.

Halliday, M. A. K., (1975). *Learning how to mean.* New York: Elsevier.

Kent, R., & Bauer, H. (1985). Vocalization of one year olds. *Journal of Child Language*, 12, 491-526.

Robb, M., & Saxman, J. (1985). Developmental trends in vocal fundamental frequency of young children. *Journal of Speech and Hearing Research*, 28, 421-427.

FOOTNOTES

Journal of Speech and Hearing Research, 33, 616-621 (1990).

†Emerson College.

††Albert Einstein College of Medicine.

†††Also City University of New York.

††††City University of New York.

Haskins Laboratories Status Report on Speech Research
1991, SR-105/106, 247-272

Patterns of Expressive Timing in Performances of a Beethoven Minuet by Nineteen Famous Pianists*

Bruno H. Repp

The quarter-note timing patterns of 19 complete performances of the third movement of Beethoven's Piano Sonata op. 31, No. 3, were measured from oscillograms and analyzed statistically. One purpose of the study was to search for a timing pattern resembling the "Beethoven pulse" (Clynes, 1983). No constant pulse was found at the surface in any of the performances. Local patterns could be interpreted as evidence for an "underlying" pulse of the kind described by Clynes, but they could also derive from structural musical factors. On the whole, the artists' timing patterns served to underline the structure of the piece; lengthening at phrase boundaries and at moments of melodic/harmonic tension were the most salient features. A principal components analysis suggested that these timing variations in the Minuet could be described in terms of two orthogonal factors, one capturing mainly phrase-final lengthening, and the other reflecting phrase-internal variation as well as tempo changes. A group of musically experienced listeners evaluated the performances on a number of rating scales. Their judgments showed some significant relations to the measured timing patterns. Principal components analysis of the rating scales yielded four dimensions interpreted as Force, Individuality, Depth, and Speed. These preliminary results are encouraging for the development of more precise methods of music performance evaluation.

INTRODUCTION

It is generally recognized that competent music performance, especially of Western art music of the past two centuries, must go beyond the written score. Without such "deviations" from the literal notation the music would sound inexpressive and mechanical, and the art of great interpreters lies largely in using such deviations with skill and taste. To the extent that the musical instrument permits it, variations in intensities, durations, and timbres of notes need to be introduced because traditional notation is not sufficiently precise in this regard as to the composer's intentions. Even those physical aspects that are precisely defined on paper—viz., fundamental frequency and timing of note onsets—require modulation by a performer to make the music come alive. Of these latter two aspects, that of variation in timing is of special interest to the psychomusicologist because it is universal to all instruments (see Gabrielsson, 1987), is a crucial aspect of performance skill, and can be measured without much difficulty.

Systematic studies of timing patterns in instrumental performance, usually on the piano, go back quite a number of years. Extensive work was done at the University of Iowa in the laboratory of Carl Seashore, who devoted a chapter in his classic book to piano performance (Seashore, 1938, pp. 225-253). Seashore and his collaborators used a photographic technique to record the hammer movements of a piano as it was played. At about the same time, Hartmann (1932) reported a detailed analysis of timing measurements derived from piano rolls. After a long hiatus during which little research of this kind seems to have been conducted, there is now renewed activity in

This research was supported by BRSG Grant RR-05596 to Haskins Laboratories. I am indebted to the record library of the Yale School of Music for providing most of the recordings analyzed here, to Larry Raphael for supplying two additional ones, and to Vin Gulisano for doing the transfers to cassette tape for me. Valuable advice in the course of this research was provided by Manfred Clynes and Robert Crowder, and helpful comments on an earlier version of the manuscript were obtained from Alf Gabrielsson and Caroline Palmer.

several laboratories, especially at the universities of Exeter (Clarke, 1982; Shaffer, 1981, 1984; Shaffer, Clarke, & Todd, 1985; Sloboda, 1983, 1985) and Uppsala (Bengtsson & Gabrielsson, 1980; Gabrielsson, 1974, 1987; Gabrielsson, Bengtsson, & Gabrielsson, 1983); see also Povel (1977) and Palmer (1989).

This research has amply confirmed the existence of systematic deviations from strict timing in the performance of experienced keyboard players, and some understanding of the rules governing the timing deviations has begun to emerge (Clarke, 1988; Sundberg, 1988; Todd, 1985). Clarke (1988) has identified three structure-governed principles within the domain of expressive timing: (1) graduated timing changes that indicate grouping of notes, with maxima at group boundaries; (2) lengthening of a note inside a group to add emphasis to the following note; (3) lengthening of structurally significant notes, especially at the beginnings of groups. Other timing rules based on local musical relationships have been proposed by Sundberg and his colleagues (see Sundberg, 1988), and composer-specific timing patterns have been postulated by Clynes (1983), based on analysis-by-synthesis techniques. Yet, both data and knowledge in this area are still quite limited, considering the diversity of musical compositions and of their possible interpretations. Quantitative analyses of music performance also tend to be laborious; for this reason, earlier studies have employed rather limited samples of music and small numbers of performers, so that their results are not necessarily representative of general principles.

The present study, though limited to a single musical composition, is the first to include a statistically representative sample of performers (N 19). Moreover, whereas most previous studies analyzed performances recorded in the laboratory, the present research follows Hartmann (1932), Povel (1977), and Gabrielsson (1987) by analyzing commercial recordings of world-famous artists. Thus the performances examined here reflect pianistic skill and interpretive insight at the highest level. The cost of this approach was some loss of measurement accuracy (partially compensated for by replication) and restriction of the investigation to timing variation only, since other measures are very difficult to obtain from sound recordings.

The goals of the study were threefold. One aim was to objectively describe and compare the expressive timing patterns of famous pianists in relation to the musical structure of the composition, to point out commonalities and individual differences, and to look for instances of expressive features observed in earlier research. A second aim was to obtain musically experienced listeners' impressions and evaluations of the various performances, to uncover the judgmental dimensions used by these listeners, and to examine whether the judgments bear any relation to the objective timing patterns. The third aim was to search for a particular timing pattern, the "Beethoven pulse," which requires some more detailed explanation.

The theory of "composers' inner pulses" was developed by Clynes (1983, 1986, 1987a), following earlier ideas by Becking (1928) and himself (Clynes, 1969). Clynes proposed that the performance of the works of the great composers of the Classical and Romantic periods, if it is to capture the composer's individual personality, requires specific patterns of timing and intensity relationships that convey the composer's individual style of movement, as it were. These living, personal pulses (distinct from a mechanically precise pulse) are said to apply to a composer's works regardless of tempo, mood, and style. The pulse pattern is assumed to be nested within hierarchical metric units and to be repeated cyclically throughout a composition. Several such composer-specific pulse patterns have been "discovered" by Clynes using computer synthesis in conjunction with his own musical judgment, and they have been implemented in computer performances of various compositions using patented software (Clynes, 1987b). Perceptual tests in which listeners of varying musical experience were presented with pieces by four different composers performed with appropriate and inappropriate pulses (Repp, 1989, 1990; Thompson, 1989) have not consistently provided support for the perceptual validity of the pulse concept, though Clynes' latest, still unpublished study did obtain positive results with subjects that included a number of outstanding musicians.

In view of these difficulties with the perceptual validation of Clynes' theory, the complementary question of whether a composer's personal pulse can be found in great artists' interpretations deserves special attention. It is noteworthy that Clynes did not rely on measurements of actual performances in deriving composers' pulse configurations; moreover, he has warned researchers that the requisite measurement accuracy cannot be achieved with current methods (Clynes, 1987a, p. 207). This warning may apply to some very subtle effects; however, many expressive deviations, such as those considered in the present study, are sufficiently large to be

measurable with reasonable accuracy even from noisy sound recordings. Although it is clear that the fixed, repetitive timing and intensity ("amplitude") patterns implemented in Clynes' computer performances are an idealization and that real performances are much more variable, Clynes' theory nevertheless implies that the pulse should be found to some extent in an excellent performance, perhaps overlaid on a multitude of structurally determined expressive deviations. Thus, for example, a performance of Beethoven's music by a pianist renowned as a Beethoven interpreter should exhibit the "Beethoven pulse" to some extent, and perhaps more so than a performance by a pianist with special expertise in the music of, say, Chopin. In addition, musical listeners' judgments of the extent to which real Beethoven performances "capture the composer's spirit" should show some positive relationship to a measure of the relative prevalence of the Beethoven pulse in these performances. These predictions were examined in the present study with respect to timing deviations.

The limitations of this enterprise should be recognized. Consideration of a single physical dimension of performance variation carries with it the danger of ignoring interactions with other dimensions, such as intensity variations. Although Clynes (1983) specifies independent timing and amplitude components of composers' pulses, this does not necessarily imply that these components are independent in actual performance; rather, they may be complementary to some extent. Another limitation is the restriction to a single composition by a single composer; clearly, a thorough search for composers' pulses in performances will eventually have to include many compositions by many composers. A third limitation, to be explained in more detail below, was that this investigation, because of the nature and tempo of the composition chosen, concerned primarily the higher level in Clynes' scheme of hierarchically nested pulses, even though the lower level (comprising time spans in the vicinity of one second) is considered the more basic one.

Despite these limitations, which diminish the impact of any negative outcome, the present study offered a valuable opportunity to provide an existence proof for Clynes' Beethoven pulse. The music chosen for this investigation was the third movement of Beethoven's Piano Sonata No. 18 in E-flat major, op. 31, No. 3, which is a representative and highly regarded work from Beethoven's early mature period. The movement has two contrasting sections (Minuet and Trio), the first having an expressive melodic line over a steady eighth-note accompaniment, and the second consisting of a kind of dialogue between "questioning" chords and "answering" melodic phrases. The Minuet, which constitutes the focus of this investigation, is well suited to a search for specific timing patterns because of its continuous movement. It might be argued that the traditional form of the Minuet imposed constraints on Beethoven's characteristic expression as well as on performers' realization of it, but this particular piece, which serves as the slow movement of the sonata, is in fact not particularly dance-like but highly expressive, at least in the Minuet section. One important consideration in choosing this music was that it was included in the perceptual tests of Repp (1989), where listeners consistently expressed a preference for a computer performance having the Beethoven pulse over performances with different pulse patterns or with none at all. Thus it seemed appropriate to search for a similar pulse in real performances of this music. The piece also offered methodological advantages: Both the Minuet and the Trio sections are divided into two parts with repeats, and the whole Minuet is repeated after the Trio, with the repeats within the Minuet again prescribed by the composer (and obeyed by most performers). Since, in addition, the two sections of the Trio are structurally very similar (if several interpolated bars are ignored) and may be treated as repetitions of each other, a single performance contains four repetitions of the musical material making up most of the composition. This fact was desirable both for a systematic assessment of performance variability across repetitions and for the reduction of measurement error by averaging across repetitions (in the absence of systematic differences).

I. MUSICAL MATERIALS

A. The composition

The third movement of Beethoven's Piano Sonata No. 18 in E-Flat major, op. 31, No. 3, is reproduced in Figure 1. It is entitled *Menuetto, Moderato e grazioso*, and has two contrasting main parts, the Minuet and the Trio. Both Minuet and Trio are in E-flat major and have 3/4 time signature. The Minuet consists of an upbeat (bar 0) followed by two 8-bar sections, labeled bars 1-8 and 9-16, respectively. Each section is repeated, with altered versions of bars 1, 8, and 16. (The two versions are labeled A and B.) There is continuous eighth-note movement in the accompaniment of the principal melody, which is rhythmically more varied and contains some sixteenth-notes.

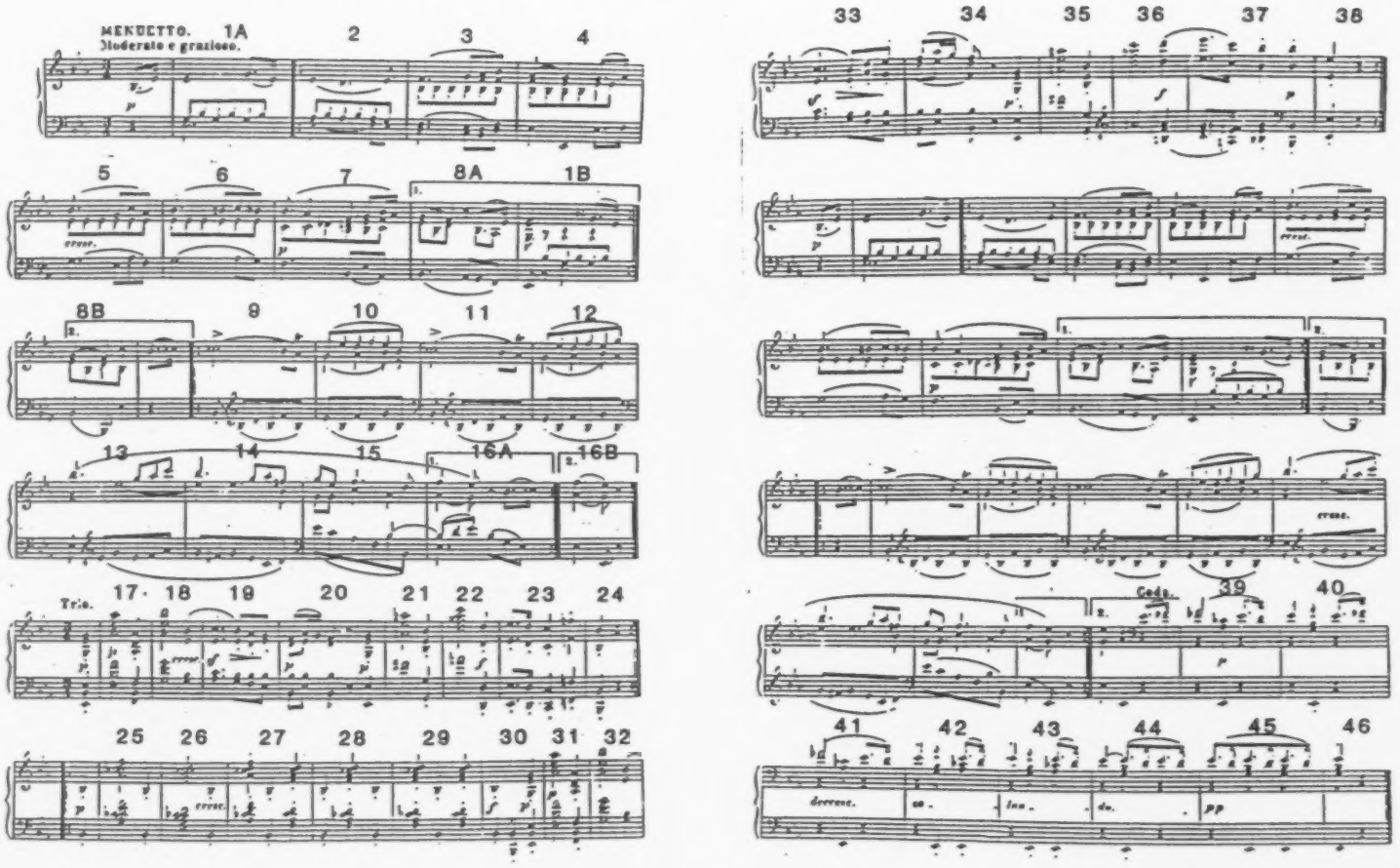



Figure 1. **The third movement of Beethoven's Piano Sonata in E-flat Major, op. 31, No. 3 (Urtext edition, Breitkopf & Härtel, 1898), with added numbering of bars.**

The Trio, too, starts with an upbeat that is followed by two sections with repeats. The first section has 8 bars (bars 17-24), while the second section has 14 bars (bars 25-38). Bars 25-30 are an interpolated ostinato passage, but bars 31-38 are very similar to bars 17-24 of the first section and were treated in the present analyses as if they were a repeat of those bars. The Trio features widely spaced, rising chord sequences followed by faster moving, falling cadences. As usual, the Minuet is repeated after the Trio; contrary to prevailing custom, however, the composer wrote the music out and prescribed repeats for each section. (It is common practice to omit section repeats in the second playing of a Minuet from the classical period, and some artists indeed disobey Beethoven's instructions in that regard.) The piece ends with an 8-bar Coda (bars 39-46) that perpetuates the rhythm of the Minuet upbeat.

For purposes of timing analysis, the piece was divided into three 8-bar sections (bars 1-8, 9-16, and 17-24/31-38), each of which occurred four times (except for those performances that omitted some repeats). The interpolated section in the Trio (bars 25-30, one repetition) and the Coda (one occurrence only) were measured but excluded from most quantitative analyses. Some analyses were conducted only on the Minuet which, because of its steady motion, was more pertinent than the Trio to the goal of detecting a continuous timing pulse.

B. The recordings

Nineteen different recordings of the Beethoven Sonata were obtained from various sources. The artists and the record labels are listed in Table 1, together with the total durations (excluding the first upbeat and the final chord) as determined by stopwatch. All except the Perahia performance (a cassette) were on regular long-playing records. Three of the performances (Davidovich, Rubinstein, Solomon) had been recorded originally from radio broadcasts onto reel-to-reel tape. Prior to measurement, all recordings were transferred to cassette tape.[1]

Table 1. *Alphabetical list of the artists and their recordings, with durations timed by stopwatch (from onset of bar 1 to onset of last bar).*

Artist	Recording	Duration
Claudio Arrau	Philips PHS 3-914	5 m 6 s
Vladimir Ashkenazy	London CS 7088	3 m 46 s
Wilhelm Backhaus	London CM 9087	3 m 43 s
Lazar Berman	Columbia M3421	4 m 26 s
Stephen Bishop	Philips 6500 392	3 m 43 s
Alfred Brendel	Vox SBVX 5418	3 m 30 s[a]
Bella Davidovich	Philips 9500 665	3 m 39 s[a]
Claude Frank	RCA VICS 9000	4 m 10 s
Walter Gieseking	Angel 35352	3 m 41 s
Emil Gilels	DG 2532 061	4 m 46 s
Glenn Gould	CBS Masterworks 7464-39547-1	4 m 12 s[a,b]
Friedrich Gulda	Orpheus OR B-1225	3 m 38 s
Clara Haskil	Epic LC 1158	4 m 4 s
Wilhelm Kempff	DG 2740 228	4 m 25 s
Murray Perahia	CBS MT 42319 (cassette)	4 m 8 s
Charles Rosen	Nonesuch NC-78010	4 m 22 s
Artur Rubinstein	RCA LM 2311	3 m 56 s[a]
Artur Schnabel	Angel GRM 4005	4 m 2 s
Solomon	EMI RLS 704 (probably)	4 m 12 s
Computer	Manfred Clynes (private cassette)	2 m 9 s[c]

[a]Section repeats not taken in second playing of Minuet.
[b]Second section repeat of Trio omitted.
[c]No section repeats at all.

In addition to the 19 human performances, a computer performance of the piece (without section repeats) was available on cassette from the earlier perceptual study (Repp, 1989). This performance had been synthesized by Manfred Clynes at the Music Research Institute of the New South Wales Conservatorium of Music in Sydney, Australia, using a special program developed there (Clynes, 1987b) to drive a Roland MKS-20 digital piano sound module. This performance instantiated the "Beethoven pulse" defined by Clynes (1983) and is described in more detail below.

II. TIMING MEASUREMENTS: QUARTER-NOTES

A. Measurement procedure

Each recording was input to a VAX 11/780 computer at a sampling rate of 10 kHz, low-pass filtered at 4.8 kHz. Portions of the digitized waveform were displayed on the large screen of a Tektronix 4010 terminal. A vertical cursor (resolution: 0.1 ms) was placed at the onsets of notes, and the times between successive cursor positions (the onset-onset interval, or OOI, durations) were recorded. If the onset of a note was difficult to determine visually, enlargement of the waveform segment on the screen sometimes helped; otherwise, the cursor was moved back in small increments and the waveform up to the cursor was played back at each step until the onset of the sought-after note was no longer audible. Only a small percentage of the measurements was obtained using this perceptual criterion, usually for accompanying notes in the initial bars of the Minuet. When several notes coincided, their individual onsets could not be resolved in the waveform, and the earliest onset was measured. This would normally have been the melody note (cf. Palmer, 1989).

Complete measurements of all recordings were made at the level of quarter-note beats (i.e., three measurements per bar). Because hand-measuring so many intervals (well over 6000) was extremely time-consuming, some accuracy was sacrificed for speed by using relatively compressed waveform displays (about 5 s per 16-inch screen). An estimate of the average measurement error was available from one recording (Ashkenazy) that was accidentally measured twice. The mean absolute discrepancy between corresponding OOI durations was 12 ms, or about 2%; the correlation was 0.98. This was considered quite satisfactory, especially since averaging over the quadruple repetitions of most of the music reduced random variability further by a factor of two. A second estimate of measurement error was obtained from the computer performance, where the two repeats of the Minuet (before and after the Trio, both measured independently) presumably were physically identical. After correcting one large mistake and omitting the values for the final note, whose duration had been deliberately extended by Clynes in synthesis, the average absolute measurement error was 6.5 ms, or less than 1%, and the correlation between the two sets of measurements was 0.99.

B. Overall tempo

The average quarter-note duration of each performance was calculated by dividing the total duration (see Table 1) by the number of quarter-note beats (348 in a performance with all repeats). From this average quarter-note duration, the average metronome speed (quarter-notes per minute, or qpm) of each performance was determined.[2] Both measures are listed in Table 2 (columns A and B), where the performances have been rearranged from slowest to fastest.

Table 2. *(A) Average quarter-note onset-onset interval durations in milliseconds. (B) The corresponding metronome speeds (quarter-notes per minute). (C) and (D) more accurate metronome speeds for the Minuet and Trio separately (see text for explanation).*

Artist	A	B	C	D
Gould	977	61	64	68
Arrau	879	68	72	71
Gilels	822	73	76	78
Rubinstein	787	76	70	89
Berman	764	79	83	82
Kempff	761	79	78	85
Rosen	753	80	85	83
Davidovich	730	82	80	87
Solomon	724	83	83	87
Frank	718	84	90	85
Perahia	713	84	85	91
Brendel	700	86	86	88
Haskil	701	86	82	91
Schnabel	695	86	86	92
Computer	694	86	84	95
Ashkenazy	649	92	95	91
Backhaus	641	94	93	105
Bishop	641	94	93	101
Gieseking	635	94	96	94
Gulda	626	96	99	92

It can be seen that there was a wide range of tempi represented, with the fastest performance (Gulda, 96 qpm) being more than 50% faster than the slowest (Gould, 61 qpm). The average metronome speed was 83 qpm. These values underestimate the underlying tempo somewhat because they include ritards, lengthenings, and pauses at phrase endings. To obtain better estimates, and also to compare the tempi for the Minuet and Trio, separate estimates for these two sections were obtained by computing the average quarter-note durations and corresponding metronome speeds from the detailed timing measurements, after excluding all bars showing conspicuous lengthening of one or more OOI durations in the grand average timing pattern (discussed below). These excluded bars were Nos. 1, 6, 7, 8, 15, 16 in the Minuet, and Nos. 24, 30, and 38 in the Trio; the Coda was also excluded. The resulting metronome speeds are shown in columns C and D in Table 2. They are indeed somewhat faster than estimated previously, with the average speed of the Minuet being 84 qpm and that of the Trio being 87 qpm.[3]

C. The three-beat Beethoven pulse

Clynes (1983) defined a composer's basic pulse as a particular pattern of time (and amplitude) relationships of the notes within a time unit of approximately 1 second. This pulse is nested within a slower, similar pulse operating on larger time units. In the Beethoven piece under investigation, the faster pulse was defined over the four sixteenth-notes within each quarter-note, whereas the slower pulse was defined over the three quarter-notes within each bar. In the computer performance, the faster pulse thus extended over a time unit of approximately 700 ms duration, whereas the slower pulse extended over about 2 seconds. The duration (and amplitude) ratios at one level are independent of those at the other.

In this study, the relative paucity of sixteenth-notes in the piece chosen led to a main focus on the slower pulse, defined over three quarter-notes within bars. The three-beat pulse used in generating the computer performance (Clynes, personal communication) had a basic timing pattern of [102.5, 94, 103.5]. This means that the first quarter-note OOI was 2.5% longer than it would have been in a mechanical performance, the second was 6% shorter, and the third was 3.5% longer.[4] Represented graphically in terms of quarter-note OOI durations, this pulse reflects a V-shaped pattern within a bar: The first and third intervals are about equally long, but the middle one is shortened.

The actual timing of the quarter-notes in the computer performance was measured in the same fashion as in the human performances. The results of these measurements are displayed in Figure 2. The upper panel of Figure 2 displays the 16 bars of the Minuet, with the two repeats superimposed; the center panel shows the Trio, with bars 31-38 laid on top of bars 17-24; and the bottom panel shows the Coda. The OOI values of the three quarter-notes within each bar have been connected to reveal the V-shaped pattern. The initial upbeat is not included; other upbeats constitute the last notes of bars 8, 16, 30, and 38, respectively. A number of bars in the Trio contain a half-note followed by a quarter-note; for these, only two values are plotted, the first of which represents half the duration of the first OOI.

The identity of the two repeats is obvious; any small discrepancies represent measurement error (see above). Large discrepancies occur in bars 16 and 24/38, where the final OOI of the Minuet preceding the Coda, the final OOI of the Trio preceding the Minuet repeat, and the upbeat of the Minuet repeat must have been deliberately extended by Clynes; another inconsistency, in bars 18/32, is of uncertain origin. A deliberate elongation of phrase-final intervals (on the second beat) is also evident in bars 8 (both repeats) and 30. Other modifications that Clynes apparently applied "by hand" to improve the musical quality of the computer performance include the prolongation of the last OOI of bar 6 (an expressive deviation that we will encounter again in many human performances), of the first OOI of bar 7, and of the second and third intervals of bar 14. The remaining bars (1-5, 9-13, 15, 16 [first repeat], 20/34, 23/37, and 27-29) exhibit the V-shaped OOI pattern characteristic of the Beethoven pulse, although some variability of the pulse shape is evident. For example, the V is deeper in bar 1 than in bars 2-5, and bars 10 and 12 exhibit an asymmetry not shared by most other bars, some of which show a smaller asymmetry in the opposite direction. This variability may represent additional adjustments made by Clynes in creating the computer performance. The two-note bars of the Trio show a rising pattern, which is consistent with the prescribed pulse, since the duration of the first value is simply the average of the first two pulse beats and thus is expected to be slightly shorter than the value for the third beat.

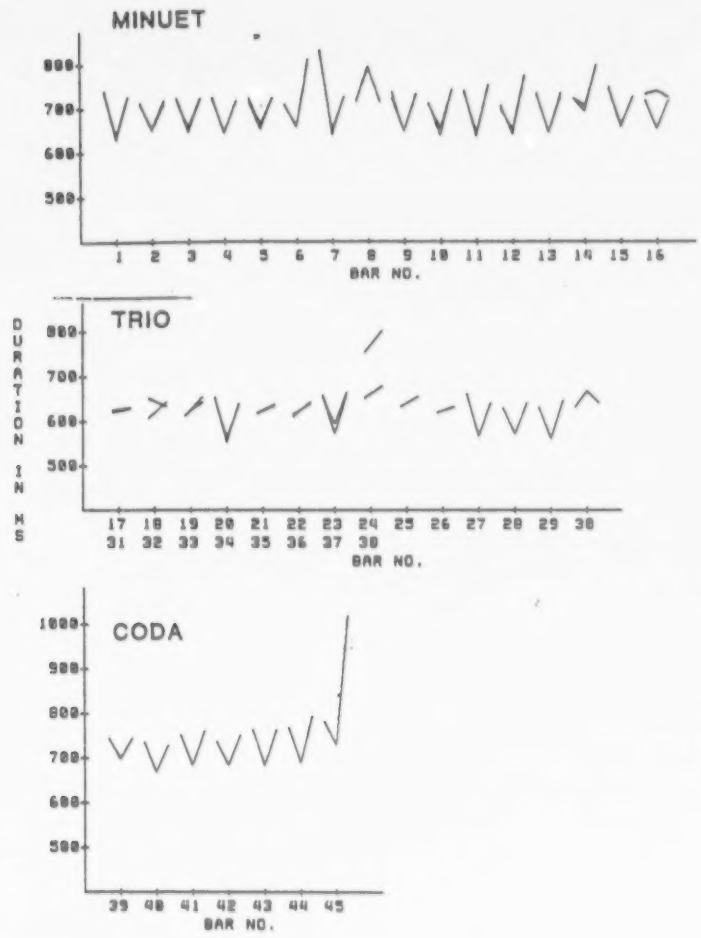


Figure 2. Quarter-note timing patterns in the computer performance created by Manfred Clynes, as measured by the author.

All in all (including repeats and Coda), there were 34 measured bars with regular V-patterns. The OOI durations in these bars were expressed as percentages of one third of the total bar duration, and the averages of these percentages were calculated across the 34 bars. The result was an average pulse of [103.7, 92.4, 103.9], which is reasonably close to (but not identical with) the pulse of [102.5, 94, 103.5] that was reportedly used in generating the computer performance.

The timing pattern of the computer performance may be considered a hypothesis about the timing pattern to be observed in expert human performances. It is not an exact prediction because human performances may be expected to be less precise and also may include additional variations in response to musical structures, that were not implemented in the computer performance. This additional variation would be superimposed on the composer's pulse (if any) to create a more complex and changing timing pattern. However, unless this additional variation is ubiquitous and large in relation to the pulse, the pulse should be detectable in the timing pattern, if it is present. Moreover, being a cyclic repetitive phenomenon, it should be present throughout a performance, though perhaps not with the consistency illustrated in Figure 2.

D. The "grand average performance"

A "grand average" timing pattern was obtained by averaging the OOI durations across 15 of the 19 human performances, keeping the repeats separate. The four performances that omitted some repeats (Brendel, Davidovich, Gould, and Rubinstein) were not included; this was just as well, since Gould's and Rubinstein's were the two most deviant performances. (See their discussion by Kaiser, 1975, pp. 340-342.) The result is plotted in Figure 3 in the format introduced by Figure 2. The lines for the four repeats are superimposed. The grand average represents those aspects of expressive timing that were common to most performances.

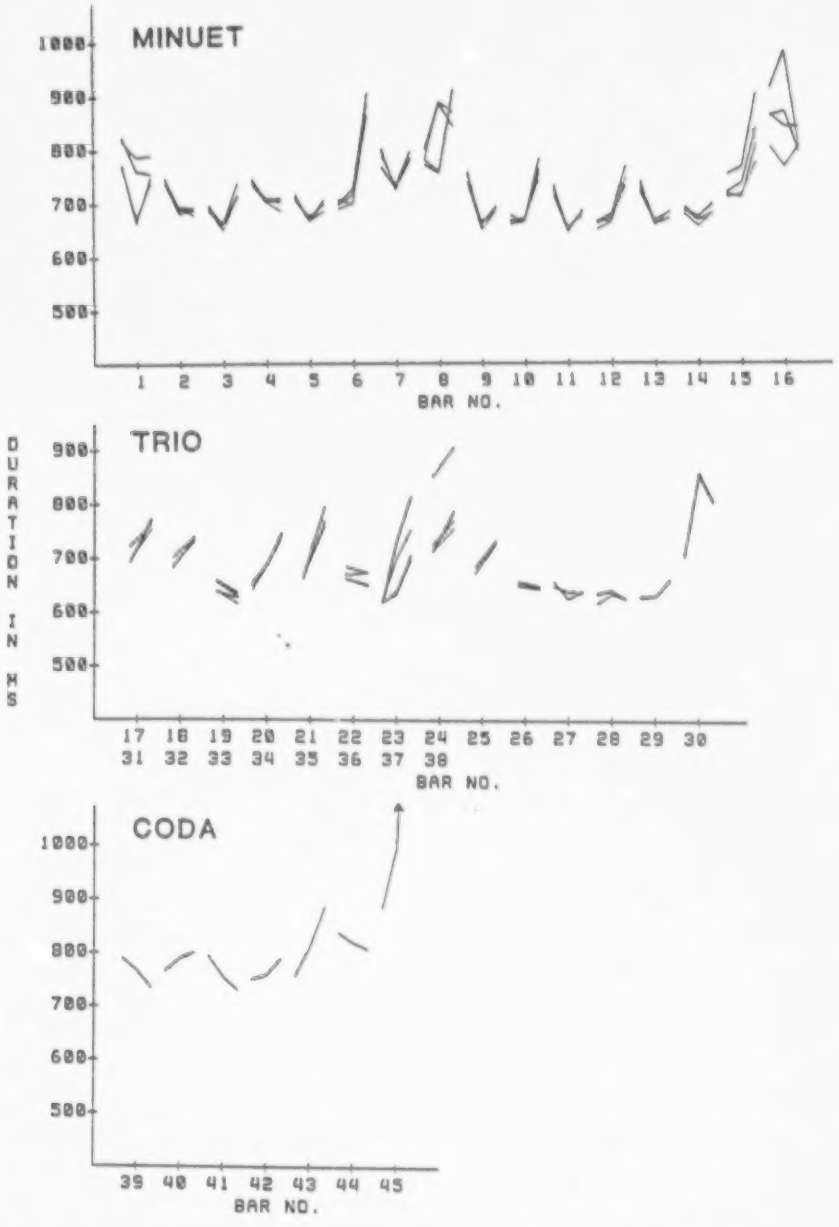


Figure 3. Grand average timing pattern of 15 human performances, with repeats plotted separately. The last data point in the Coda (arrow) is at 1538 ms.

1. Repeats

It is evident, first, that repeats of the same material had extremely similar timing patterns. This consistency of professional keyboard players with respect to detailed timing patterns has been noted many times in the literature, beginning with Seashore (1938, p. 244). The only systematic deviations occurred in bar 1 and at phrase endings (bars 8, 15-16, and 23/37-24/38), where the music was in fact not identical across repeats (see Figure 1): In bar 1 Beethoven added an ornament (a turn on E-flat) in the repeat (bar 1B), which was slightly drawn out by most pianists. In bar 8A, which led back to the beginning of the Minuet, the upbeat was prolonged, but in bar 8B, which led into the second section of the Minuet, an additional ritard occurred on the phrase-final (second) beat. Similarly, a uniform ritard was produced in bar 16A, which led back to the beginning of the second Minuet section, and an even stronger ritard occurred on the phrase-final (first and second) notes of bar 16B, which constituted the end of the Minuet, whereas the third note constituted the upbeat to the Trio and was taken shorter. Bar 15 anticipated these changes, which were more pronounced in the second playing of the Minuet, following the Trio. Similarly, bar 37 anticipated the large ritard in bar 38 when, in the second repeat, it concluded the Trio. Another phrase-final phenomenon, consistent across both repeats, occurred in bar 30 of the Trio, at the end of the interpolated section. All these deviations confirm the well-known principle of phrase-final lengthening (e.g., Clarke, 1988; Lindblom, 1978; Todd, 1985). Only one major phrase-internal expressive deviation was evident, the lengthening of the third beat in bar 6. That beat not only carries an important melodic inflection in eighth-notes, but also is followed by an abrupt change in dynamic level (*subito piano*) in the score (see Figure 1). Thus it is an example of Clarke's (1988) second and third principles: lengthening of a note to underline its own structural significance as well as to enhance the next note (which, being an appoggiatura, needs the enhancement especially because of the sudden reduction in intensity).

The consistency of the grand average timing pattern across repeats in the Minuet was quantified by computing the correlations of OOI values between (a) immediate repeats (where slight changes in the music occurred) and (b) distant repeats (i.e., of the identical music before and after the Trio). The correlations were computed separately for the two Minuet sections and then averaged. These correlations were (a) 0.85 and (b) 0.95, the former being predictably lower because of the musical and interpretive changes discussed above. Corresponding correlations were also calculated for each of the 19 individual performances, to determine the individual consistency of the different artists. The average individual correlations were (a) 0.66 and (b) 0.79. Only one pianist (Davidovich) showed exceptionally low correlations across repeats (0.36, 0.32); all others showed rather high consistency. The correlations are lower than those for the grand average because they include random as well as perhaps intended but idiosyncratic variation across repeats that cancelled out in the grand average.

It can be seen in Figure 3 that the performances were similarly consistent across repeats in the Trio.

2. Pulse-like patterns

Consider now the within-bar OOI patterns in the grand average. According to Clynes' theory, they should follow a fairly consistent V-shaped pattern, especially in those bars that do not contain major deviations due to phrase boundaries or special emphasis. It is clear from Figure 3 that in the Minuet only two bars show the nearly symmetric V-shape of the Beethoven pulse: bar 1 (first repeat only) and bar 7. Bar 5 shows a shallower V-shape, bar 3 has an asymmetric V-shape, and so do bars 9, 11, and 13, but with the asymmetry going in the opposite direction. Interestingly, these bars are all odd-numbered ones. The even-numbered bars, discounting those with major expressive deviations (6, 8, 16), do not show V-shapes: Bars 2 and 4 show only an elongated first OOI, bars 10 and 12 an elongated third OOI, and bar 14 is almost evenly timed.

This curious alternating pattern makes good sense when the melodic and harmonic structure of the Minuet is considered. Moments of tension (relative dissonance, elevated pitch) alternate with relaxation (high consonance, lower pitch). The former occur on the first beats of bars 2, 4, 7, 9, 11, 13, and 16, which are all prolonged; the latter fall on the following beat, which is invariably shortened. The beat preceding a moment of high tension also tends to be prolonged (third beats of bars 1, 3, 6, 8, 10, 12, and 15). Thus it seems that the timing pattern in the Minuet was determined primarily by the expressive requirements of the melody, not by a constant, autonomous pulse.

The results for the Trio and for the Coda reinforce these conclusions. None of the bars with three measured OOI values shows a V-like shape:

Some shapes are rising, others are flat or falling. Of the bars with only two measurements, some show a rising pattern, but others a slightly falling one. The observed patterns, however, again make sense with respect to the musical events. Basically, the pianists tended to take the staccato upbeats preceding the half-notes somewhat longer than notated, perhaps to create a slight suspense, or simply as a physical consequence of the large leap upwards. Other bars in the Trio, including the ostinato in the interpolated section (bars 26-29), were played with very even timing. In the Coda, a tension-relaxation pattern is evident, which follows the melodic contour and echoes that observed in bars 9-12 of the Minuet.

Although the grand average does not show any evidence of a *constant* V-shaped pulse, Figure 3 does suggest an overall tendency for the second OOI in a bar to be shorter than the first and the third. To examine whether these within-bar differences in relative OOI duration were consistent across pianists, a repeated-measures analysis of variance (ANOVA) was conducted on the Minuet OOI data after converting them to percentage values of the total bar durations (as used in Clynes' pulse specifications), which effectively eliminated tempo differences among different performances, and among bars within performances. The analysis included only the 15 performances with complete repeats, and only the 11 bars (2-5, 7, and 9-14) without major expressive deviations and/or variations across repeats. The fixed factors in the ANOVA were Beats (3), Bars (11), and Repeats (4); the random factor was Pianists (15). There was indeed a highly significant main effect of Beats [$F(2,28) = 43.92$, $p < .0001$], which confirms that, overall, there were reliable differences among the three OOI values in a bar. The average percentages were [102.2, 96.3, 101.5], showing the predicted reduction of the second OOI. However, there was also a highly significant interaction of Beats with Bars [$F(20,280) = 17.96$, $p < .0001$], which shows the bar-to-bar variations in OOI patterns (cf. Figure 3) to be reliable across pianists. While the overall "pulse" (the Beats main effect) was highly significant with Pianists as the sampling variable, it was less reliable, though still significant, when Bars were considered the sampling variable—that is, when the size of the Beats main effect was compared to that of the Beats by Bars interaction [$F(2,20) = 7.22$, $p < .01$]. These results permit the interpretation that there *is* some underlying constant pulse that, together with local musical requirements, contributes to the surface pattern of OOI durations. To the extent that the average pattern of [102.2, 96.3, 101.5] is the best estimate of such an underlying pulse (which it may not be), it is not radically different from the pattern applied by Clynes in the computer performance, [102.5, 94, 103.5], though a separate analysis showed the difference to be significant, [$F(2,20) = 19.06$, $p < .0001$].

E. Differences in timing patterns among individual pianists

Even though the grand average timing pattern did not show much evidence of a continuous V-shaped pulse, the possibility exists that certain individual pianists did exhibit such a pattern to a greater extent. The discovery of a composer's personal pulse is said by Clynes (1987a) to be restricted to those who are intimately familiar with a composer. Although all of the 19 great artists examined here must be (have been) thoroughly familiar with Beethoven's works, some of them are nevertheless considered greater Beethoven interpreters than others by critics and concert audiences, and they also differ in how often they perform(ed) Beethoven's music. In the overall ANOVA, when Repeats (rather than Pianists) were considered the random factor, there was in fact a significant Pianists by Beats interaction [$F(28,84) = 10.16$, $p < .0001$], showing that the artists differed in their average within-bar timing patterns. There was also a significant Pianists by Bars by Beats interaction [$F(280,840) = 4.40$, $p < .0001$], indicating that the artists varied their timing patterns across bars in different ways.

Individual analyses of variance were conducted on each pianist's Minuet data, again expressing OOI durations as percentages within bars (i.e., eliminating tempo variations across bars and across pianists) and including only the 11 bars without major expressive deviations. Repeats served as the random factor in these analyses, its interaction terms providing the error estimates. The results of these analyses are summarized in Table 3. The average OOI patterns of the individual artists show some striking similarities: All artists but one (Ashkenazy) prolonged the first OOI somewhat, and all, without exception, reduced the second OOI by varying amounts. The majority did not change the last OOI much, though some prolonged it. These consistencies explain the statistical reliability across pianists of the grand average OOI pattern described above. The individual average OOI patterns were reliably different from mechanical evenness ([100, 100,

100]) *across repeats* for all pianists but one (Gould), though some pianists showed more consistently expressive patterns than others. (There appears to be no relation to the artists' renown as Beethoven interpreters.) Only eight pianists, however, showed a reliable OOI pattern *across bars*, and then only at the $p < .01$ level. Moreover, all pianists showed a highly significant ($p < .0001$) Beats by Bars interaction; that is, every one of them (even Gould) varied the timing pattern between bars and maintained these variations systematically across repeats. Note that this analysis concerned only those bars that did not have any major expressive deviations to begin with. Thus, no single artist showed any constant pulse at the surface, though eight of them might be credited with a possible underlying pulse that was overlaid by expressive variations of a different kind.

Table 3. *Individual pianists' average timing patterns (in onset-onset interval percentages) across 11 bars of the Minuet; significance levels of the Beats main effect across repeats (R) and across bars (B); and mean squares term of the Beats by Bars interaction (MSQ), which provides a measure of the relative bar-to-bar variability of the timing pattern. Key: * p .0001, * p .001, p .01.*

Artist OOI	pattern	R	B	MSQ
Arrau	[101.2, 96.0,102.8]	**		161
Ashkenazy	[98.9, 97.0,104.1]	**	*	89
Backhaus	[104.1, 97.2, 98.7]	**	*	68
Berman	[101.6, 92.7,105.7]	***	*	241
Bishop	[102.5, 95.5,102.0]	**		183
Brendel	[102.9, 96.2,100.9]	*		89
Davidovich	[102.5, 96.1,101.4]	**		68
Frank	[101.9, 97.2,100.9]	**		109
Gieseking	[103.1, 95.3,101.6]	**	*	76
Gilels	[101.8, 95.7,102.5]	*	*	67
Gould	[101.0, 99.1, 99.9]			21
Gulda	[102.7, 97.3,100.0]	**	*	46
Haskil	[101.8, 98.7, 99.5]	**		66
Kempff	[103.4, 96.5,100.1]	***	*	79
Perahia	[103.1, 95.4,101.5]	***	*	99
Rosen	[102.6, 94.7,102.7]	**		202
Rubinstein	[102.1, 95.2,102.7]	**		259
Schnabel	[101.3, 97.9,100.8]	*		117
Solomon	[103.6, 96.8, 99.6]	*		124

The last column in Table 3 lists the mean square terms of the Beats by Bars interaction, which provide a relative numerical measure of how much the OOI pattern varied from bar to bar. *These values are correlated with the "depth" of* the average OOI modulations: Pianists with a shallow average pattern (e.g., Gould, Haskil, Gulda) also tended to vary less from bar to bar, whereas pianists with a highly modulated average pattern (e.g., Berman, Rosen, Rubinstein) also showed large variations from bar to bar. In other words, the highly expressive pianists in the latter group, whose average timing patterns resembled most that of the postulated Beethoven pulse, were least inclined to maintain a constant pulse throughout. The V-shaped average pattern may well be the consequence of structural musical factors that, on the whole, favored relative lengthening of the first and third beats, rather than the reflection of an underlying autonomous Beethoven pulse.

F. Factor analysis of timing patterns

So far, the analysis has considered only a subset of the bars in the Minuet, those without major expressive deviations, and tempo variations across bars have been disregarded. A more comprehensive analysis was conducted on the complete Minuet quarter-note OOI data (absolute durations, averaged over repeats) of all 19 pianists. The statistical technique employed was principal components factor analysis with Varimax rotation, which reveals the structure in the matrix of intercorrelations among pianists' timing patterns. One purpose of the analysis was to determine whether the individual differences in timing patterns could be described in terms of a single factor (implying that all pianists follow the same pattern, only in different degrees), or whether several factors would emerge. The second purpose was to see whether a factor could be extracted that reflects the hypothetical underlying Beethoven pulse. If there is such a pulse that combines additively with timing variations of a different origin, then principal components analysis would seem to be a good method of separating these different sources of variation.

To facilitate extraction of a "pulse" factor, the computer performance, which instantiated the Beethoven pulse, was included in the analysis. This computation on the 20 × 20 intercorrelation matrix yielded three orthogonal factors considered significant by the criterion that their eigenvalues were greater than one. These three factors together accounted for 77% of the variance in the data. Before rotation, the first factor accounted for most of that variance (63%), with the other two factors adding only 8% and 6% of variance explained, respectively. After Varimax rotation,

which aims for a "simple pattern" of factor loadings, the variance accounted for was more evenly divided among the three factors: 31.2%, 26.6%, and 19.0%, respectively. Table 4 shows the factor loadings of the individual performances (i.e., the correlations of individual timing patterns with the patterns characterizing each of the three factors) and their communalities (i.e., their squared multiple correlations with all three factors, which represent the variance explained by the factors). Figure 4 shows the factor timing patterns themselves, rescaled into the millisecond domain by multiplying the standardized factor scores with the average standard deviation and adding this product to the grand mean. In interpreting these data, it should be kept in mind that all the performances, with few exceptions, showed substantial positive intercorrelations (0.5-0.8) which were caused by the major expressive excursions and ritards shared by most artists.

Table 4. *Factor loadings and communalities (squared multiple correlations) of the various performances in the three rotated factors found by principal component analysis of the Minuet data.*

	Factor 1	Factor 2	Factor 3	Communality
Arrau	0.662	0.371	0.341	0.692
Ashkenazy	0.380	0.666	0.335	0.700
Backhaus	0.085	0.884	0.214	0.834
Berman	0.579	0.352	0.630	0.865
Bishop	0.333	0.623	0.426	0.681
Brendel	0.261	0.360	0.749	0.758
Davidovich	0.372	0.517	0.509	0.664
Frank	0.665	0.581	0.289	0.863
Gieseking	0.181	0.704	0.403	0.692
Gilels	0.707	0.506	0.332	0.865
Gould	0.919	-0.210	0.181	0.921
Gulda	0.569	0.307	0.613	0.793
Haskil	0.567	0.539	-0.056	0.616
Kempff	0.284	0.618	0.491	0.704
Perahia	0.807	0.395	0.293	0.893
Rosen	0.691	0.450	0.404	0.843
Rubinstein	0.768	0.407	0.228	0.807
Schnabel	0.582	0.644	0.122	0.769
Solomon	0.615	0.452	0.385	0.731
Computer	0.104	0.111	0.807	0.674

The first factor represents primarily the phrase-final lengthenings in bars 8 and 15-16. The timing pattern in bars 9-13 is weakly represented, as is the phrase-final lengthening in bar 4. Most artists have high loadings on this factor, which means that they dutifully marked the major phrase boundaries. The highest correlation is exhibited by Gould (whose performance offered little else), followed by Perahia and Rubinstein; low correlations are shown by Backhaus, Clynes' computer performance, and Gieseking.

The second factor, orthogonal to the first, also represents the phrase-final lengthening in bars 8 and 15-16, though less strongly, and in addition shows a "slow start," the expressive lengthening in bar 6, and a strong tendency for the second half of the Minuet to be faster than the first. This last feature was especially obvious in Backhaus' performance, which also shows the highest loading on this factor, followed by Gieseking and Ashkenazy. Low loadings are exhibited by Gould and Clynes' computer performance.

The third factor, in striking contrast to the other two factors, represents a relatively even, V-shaped timing pattern, though its depth exceeds that of the pulse implemented in the computer performance (cf. Figure 2). Not surprisingly, the computer performance exhibits the highest loading on this factor, followed by Brendel, Berman, and Gulda. Low loadings are associated with Haskil, Gould, and Schnabel.

The emergence of this third factor is intriguing and might be taken as a confirmation of an underlying Beethoven pulse in at least some of the performances. However, this factor was due to the inclusion of the computer performance in the analysis: When the analysis was repeated with the computer performance excluded, it returned only two factors that together accounted for 73% of the variance, which was split about evenly after rotation. The first factor was quite similar to that of the previous analysis, while the second factor was a conflation of the second and third factors obtained earlier; that is, it exhibited various tilted V-shaped patterns within bars, similar to those seen in the grand average performance (Figure 3). It is not clear, therefore, how much importance should be attached to the extraction of a separate pulse factor when the computer performance was included. Nevertheless, that analysis provides a description of the individual performances in terms of several independent timing aspects, and it successfully isolates a pulse-like aspect from other component patterns. The loadings in Factor 3 (Table 4) provide a measure of the degree of presence of an underlying pulse in individual performances, regardless of whether or not the evidence is deemed sufficient for concluding that there is such a pulse. We will examine later whether this measure shows any relation to judgments of the performances as more or less "Beethovenian."

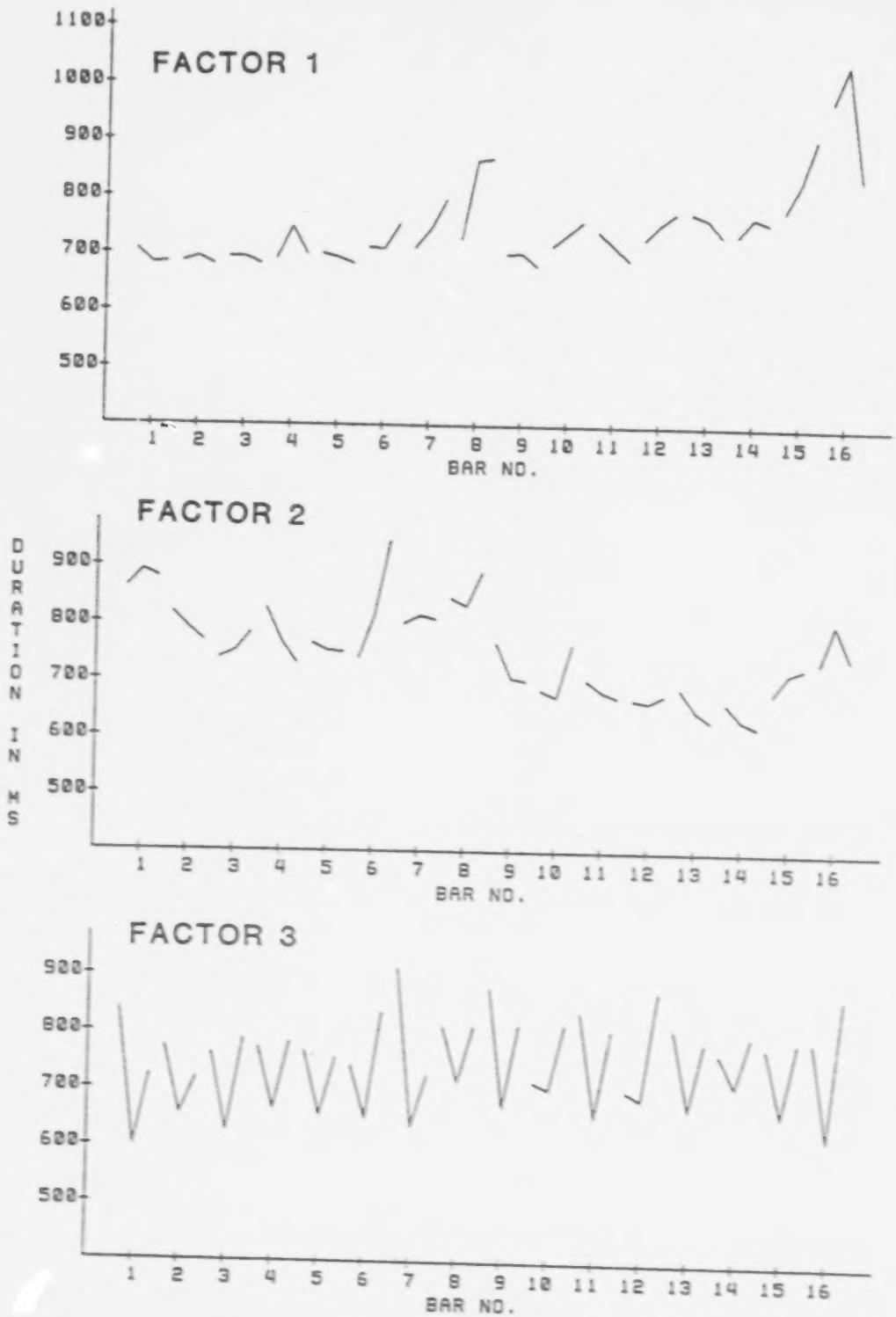


Figure 4. **Factor timing patterns emerging from the principal components analysis including the computer performance.**

III. TIMING MEASUREMENTS: SIXTEENTH- AND EIGHTH-NOTES

The analyses so far have concerned the temporal microstructure at the level of quarter-notes, which constitutes the higher level in Clynes' pulse hierarchy for this particular piece. This level was relatively easy to access and measure. Within each quarter-note, however, Clynes defined a four-beat Beethoven pulse, which forms the lower level in the pulse hierarchy. This level was more difficult to evaluate because its full expression required sixteenth-notes, which were quite rare in the present composition. Eighth-notes were common but provided less information, since they reduced the four-beat pulse to a two-beat pulse. Some measurement problems were also encountered. Nevertheless, some data were obtained about the temporal microstructure at this level.

A. Sixteenth-notes

1. Measurement procedures

Sequences of two sixteenth-notes occur in several places (bars 7, 20, and 34), but proved very difficult to measure; the onset of the second note could usually not be found in the acoustic waveform. Therefore, the measurements were restricted to single sixteenth-notes following a dotted eighth-note. Such notes occur in bars 0/8A, 1, 4, and 8B/16A of the Minuet, in bar 23/37 of the Trio, and throughout the Coda. With four repeats of the Minuet and two of the Trio in most performances, there were generally four independent measures available for each of the four sixteenth-note occurrences in the Minuet and for the single occurrence in the Trio (the latter really being two similar occurrences, each repeated twice). For the Coda, of course, only a single set of measurements was available for each artist, but there were 11 occurrences of sixteenth-notes.

The measurements were performed as previously, but with less of the waveform (about 2 s) displayed at a time, to reduce measurement error. In each of the relevant bars, the OOI durations of the dotted eighth-note and the following sixteenth-note were measured. The relative duration of the sixteenth-note OOI following a dotted eighth-note was computed as the ratio of its measured duration to its "expected" duration, times 100. The expected duration was one fourth of the sum of the dotted eight-note and sixteenth-note OOI durations.

2. The four-beat Beethoven pulse

Clynes' four-beat Beethoven pulse is defined as [106, 89, 96, 111] (Clynes, 1983). Thus it implies that a sixteenth-note following a dotted eighth-note should be 11 percent longer than its expected value, and that this difference (with due allowance for random variability and other expressive demands of the piece) should be present throughout a Beethoven composition. The aim of the present measurements was to examine these predictions.

To rule out any possible misunderstanding, the computer performance synthesized by Clynes was measured to determine the relative lengths of the sixteenth-notes in it. Since that performance did not contain any repeats, only two independent measurements (on identical renditions) were available for the four occurrences of sixteenth-notes in the Minuet and the one in the Trio. These two measurements were in close agreement; the average absolute difference due to measurement error was 1.2 percentage points. The average relative OOI durations of the five sixteenth-notes were: 107.5, 108, 103.5, 110.5, 111.5. The last two are in agreement with Clynes' specifications, but the first three are smaller, the third (bar 4) conspicuously so. A similar impression was obtained from the Coda; the 11 values were: 106, 102, 110, 112, 99, 111, 110, 108, 115, 117, 147. The first, second, and fifth values are clearly smaller than expected, and the last three are larger, though the last one obviously includes a manual adjustment, the final ritard. The origin of this variability in Clynes' own creation remains unclear. The measurements nevertheless illustrate the postulated lengthening of sixteenth-notes following dotted eighth-notes.

3. Results

The relative durations of the sixteenth-notes in the 19 human performances are presented in Table 5. Each percentage value for bars 0/8A, 1, 4, 8B/16A, and 23/37 represents an average across four (sometimes fewer) repeats; for the Coda, a single average was computed from the eight values of bars 16B–44, but separate values are reported for the three occurrences in the final bar (45a, b, c).

Consider first the four occurrences in the Minuet (the first four columns in Table 5). There are very large individual differences among the artists, particularly in the initial upbeat, bar 0/8A. Some pianists (Ashkenazy, Berman, Rosen, Frank) show a considerable elongation of the sixteenth-note in that beat, while Schnabel, the great Beethoven authority, played it much shorter than notated. Many others are close to precise timing.

Table 5. *Relative OOI durations of sixteenth-notes following dotted eighth-notes (in percent).*

	Minuet				Trio		Coda		
	Bar 0/8A	1	4	8B/16A	23/37	16B-44	45a	45b	45c
Arrau	109.5	120.3	111.2	119.0	87.8	89.4	100.0	104.0	144.0
Ashkenazy	145.8	120.8	108.3	126.8	85.8	98.8	107.0	104.0	185.0
Backhaus	101.3	92.8	84.0	94.0	83.3	85.5	87.0	110.0	152.0
Berman	129.8	126.5	110.8	131.8	84.8	87.5	84.0	88.0	127.0
Bishop	101.8	111.3	119.8	108.0	82.0	86.6	108.0	123.0	143.0
Brendel	103.8	90.5	124.8	94.8	79.3	90.6	109.0	115.0	125.0
Davidovich	98.5	102.0	108.3	100.3	85.5	92.1	111.0	104.0	116.0
Frank	125.8	112.5	103.0	123.5	86.5	77.8	86.0	101.0	168.0
Gieseking	101.8	101.3	96.5	96.5	101.3	100.1	113.0	113.0	142.0
Gilels	94.5	104.8	105.3	85.5	81.5	80.6	92.0	106.0	130.0
Gould	105.5	112.8	100.5	102.8	81.0	70.6	77.0	84.0	112.0
Gulda	98.3	106.0	94.8	107.3	84.5	93.9	118.0	106.0	164.0
Haskil	108.5	107.8	114.5	92.5	93.3	87.1	86.0	84.0	133.0
Kempff	98.8	100.8	110.0	108.8	81.8	83.4	94.0	98.0	123.0
Perahia	115.5	120.8	126.8	103.0	92.8	97.8	127.0	134.0	199.0
Rosen	129.0	142.3	134.3	121.8	100.3	86.5	96.0	96.0	119.0
Rubinstein	102.0	104.5	107.0	96.5	79.8	87.9	95.0	100.0	149.0
Schnabel	78.3	94.0	99.5	89.3	90.5	78.8	60.0	69.0	116.0
Solomon	101.5	122.0	108.8	105.0	80.8	89.8	90.0	88.0	135.0
Computer	107.5	108.0	103.5	110.5	111.5	107.3	115.0	117.0	147.0

Though not shown in the table, a number of artists (most notably Ashkenazy, Backhaus, Frank, Kempff, Perahia, and Rosen) prolonged the sixteenth-note much more in the repeat (bar 8A) than in the first playing (bar 0). Despite considerable variability, individual differences tended to be maintained throughout the Minuet. Some of the variability, of course, may have been intended by the artist. It is perhaps noteworthy that the pianists with rather short sixteenth-notes are mostly of German extraction, while those with prolonged sixteenth-notes are Russian or American.

The Minuet data were subjected to a repeated-measures analysis of variance with two fixed factors, Bars and Repeats.[5] The main effect of Bars was not significant, showing that the artists as a group did not systematically modulate their characteristic sixteenth-note timing patterns within the Minuet. The main effect of Repeats was marginally significant, $F(3,54) = 3.20$, $p = 0.0304$, due to a tendency for longer sixteenth-notes in the second repeat. However, this tendency occurred only in bars 0/8A and 8B/16A, whereas bars 1 and 4 showed the opposite pattern, with longer sixteenth-notes on the first playing. This difference was reflected in a significant Bars by Repeats interaction, $F(9,162) = 2.63$, $p = 0.0074$.

On the whole, despite all the individual variations, there was a tendency to prolong the sixteenth-notes in the Minuet, as predicted by Clynes' Beethoven pulse. The grand average percentage deviation was 108.1, which is not too far from Clynes' 111. The situation is very different, however, for the Trio and the Coda. In the Trio (bar 23/37), there was an overwhelming trend to shorten the sixteenth-note. Only two pianists (Gieseking, Rosen) played the note with its literal value. The average percentage value was 86.4, which strongly contradicts the value of 111 implemented in Clynes' synthesis. Similarly, throughout the Coda (bars 16B-44), except for the last bar, there was a strong tendency to shorten all sixteenth-notes. A few pianists (Gieseking, Ashkenazy, Perahia) played almost literally, but none showed any lengthening. The average percentage value was 87.6, which again contrasts

with Clynes' nominal 111. The first four occurrences tended to be somewhat longer than the second four, which was reflected in a significant difference among the eight values in bars 16B-44, $F(7,126) = 2.73$, $p = 0.0114$. In the final bar (bar 45), some pianists started to lengthen the relative duration of the sixteenth-notes in the first two beats, and the very last occurrence (45c) showed a substantial lengthening in all performances, due to the final ritard. However, the average percentage values for the first two occurrences in bar 45, 96.8 and 101.4, are still far below Clynes' nominal 111, while the last occurrence, with an average of 141.2, comes close to Clynes' deliberately extended 147.0.

In summary, these results offer some support for Clynes' specifications during the Minuet, although there were large individual differences, and the most deviant pianists included some of the most renowned Beethoven players (Schnabel, Backhaus, Brendel). The Trio and Coda performances, however, flatly contradict Clynes' Beethoven pulse. It appears that different musical requirements call for radically different temporal microstructures, even for the same composer and within the same piece.

B. Eighth-notes

1. Measurement procedure

Eighth-notes are present throughout the Minuet and form a continuous background movement against which the melody unfolds. However, the onsets of single accompanying eighth-notes were often very difficult to detect in the background noise, and complete measurements did not promise enough information to seem worth the effort. Therefore, eighth-notes were measured only in two selected bars, bars 10 and 12, where they formed part of the melody (cf. Figure 1). Each of these bars contains three pairs of eighth-notes, with usually four repeats per performance. The relative OOI durations of each pair were expressed simply as their ratio, with the second OOI in the numerator.

2. The two-beat Beethoven pulse

For two eighth-notes, the pulse is reduced to a two-beat pattern, [97, 103].[6] Thus, the second note in any pair of eighth-notes nested within a quarter-note beat is predicted to be 3 percent longer than expected, and the ratio between the two OOIs should be 103/97 1.06.

The three pairs of OOI durations were measured very carefully in bars 10 and 12 of the computer performance, and their ratios were computed. They were 1.11, 1.12, and 1.17 in bar 10; and 1.08, 1.06, and 1.24 in bar 12. It is evident that, in each bar, the last ratio was considerably larger than prescribed by the pulse specification. The first two ratios in bar 10 also appear too large. The cause of these deviations is unknown; they may represent further personal interventions by Clynes to improve the quality of the computer performance beyond that imparted by the constant pulse pattern.

3. Results

The results for the human performances are easily summarized. The ratio data for the 15 pianists who observed all the repeats were subjected to a repeated-measures analysis of variance with the factors Distant Repeats (before versus after the Minuet), Immediate Repeats, Bars (bar 10 versus bar 12), and Beats (1, 2, 3). There was a single, highly significant effect: the main effect of Beats, $F(2,28) = 56.03$, $p < .0001$. It was due to an extremely consistent tendency to lengthen the second eighth-note on the third beat, but not on the first and second beats. The three average ratios were: 1.01, 1.01, and 1.18. None of these ratios matches the value of 1.06 predicted by Clynes' specifications; in particular, the first two beats do not show the predicted relative lengthening of the second eighth-note. The substantial lengthening of the last eighth-note in each bar is in agreement with the pattern implemented in the computer performance, but not with the Beethoven pulse.

The consistency of the observed pattern across pianists, including the four not included in the analysis of variance, was striking. Only two pianists did not show the increased OOI ratio on the third quarter-note beat: Solomon (who played close to mechanical evenness) and Gould (who tended towards slightly positive ratios on all three beats, not unlike the predicted Beethoven pulse). A few pianists (Schnabel, Brendel, Ashkenazy, Bishop) tended to lengthen the second eighth-note OOI on the first beat as well, and some (Berman, Gulda, Ashkenazy) showed some lengthening of the second eighth-note on the second beat, while one (Schnabel) showed some shortening.

On the whole, the eighth-note data show once more that timing relationships are not constant but vary according to local musical requirements. The musical factors that caused lengthening of the last eighth-notes in bars 10 and 12 are phrase-finality (as indicated by the slurs in the score) and, perhaps more importantly, emphasis of the following note (Clarke's, 1988, principles 1 and 2).

IV. PERFORMANCE EVALUATION BY LISTENERS

This part of the study had two purposes: First, to get musical listeners' impressions of the "Beethovenian" quality of the performances and to see whether these ratings bore any relationship to the timing measurements, particularly the "Beethoven pulse" factor (Factor 3) in the principal components analysis. Second, to explore in a very preliminary way the psychological dimensions along which musical performances vary. This part of the study was exploratory in nature, and its methodology did not include all the controls one would want to include in a full-scale study of human performance evaluation. Nevertheless, it yielded some interpretable results.

A. Subjects

Nine subjects participated, all of whom had extensive experience with classical music, played the piano with varying degrees of proficiency, and knew the piece well. They included a senior professor of piano at a major music school; a young professional pianist who frequently concertized in the area; an experienced piano teacher at a community music school; two doctoral students of musicology with a special interest in performance; a distinguished professor of phonetics with a long-standing interest in rhythm; an old lady (the author's mother) with a life-long interest in music performance; and two psychologists with a strong interest in music (the author being one of them). Four of the listeners were of European origin (two Austrians, one German, one Estonian); the others were American. Four judges were female; five were male.

B. Procedure

Twenty adjective pairs were selected by the author with the intention of capturing dimensions relevant to the judgment of performance variations. They are shown in Table 6.[7] A seven-point rating scale extended between each pair of polar opposites. The most important adjective pair for the present purpose was the first one (Beethovenian/Un-Beethovenian), which was explained in the instructions as follows: "This judgment should reflect to what extent the performance 'captures the Beethoven spirit'—that is, whether the general expressive quality of the performance is what the composer may have intended." The second adjective pair concerned the perceived tempo of the performance. The remaining adjective pairs included some deliberately selected because they were thought to correlate with the "Beethovenian" quality; for example, Clynes (1983) characterizes the Beethoven pulse as "restrained." The order of the adjective pairs and the assignment of their members to the ends of the rating scale were fixed in this exploratory study. The lower end of the rating scale represented the more positive connotation for some adjective pairs, the more negative one for others.

Table 6. *The adjective pairs used in performance evaluation.*

Beethovenian	Un-Beethovenian
Fast	Slow
Expressive	Inexpressive
Relaxed	Tense
Superficial	Deep
Cold	Warm
Powerful	Weak
Serious	Playful
Pessimistic	Optimistic
Smooth	Rough
Spontaneous	Deliberate
Consistent	Variable
Coherent	Incoherent
Sloppy	Precise
Excessive	Restrained
Rigid	Flexible
Effortful	Facile
Soft	Hard
Realistic	Idealistic
Usual	Unusual

Each subject was given a booklet containing twenty identical answer sheets, one for each performance, preceded by detailed instructions that included brief definitions of each adjective pair. The instructions emphasized that the musical, not the sonic quality of each performance was to be judged. They gave the option of entering two separate judgments for Minuet and Trio on each scale; that option was rarely taken, and if it was, the two judgments were averaged in the following data analysis. A few missing judgments were replaced by "4."

All subjects listened at home on their own audio equipment to a cassette tape of the 20 performances, stopping the tape after each performance to enter their judgments. The 20 performances were in the same fixed order for each subject. The first performance was the computer rendition, both to get an immediate reaction to it and to remind the subjects of the music. The order of the human performances was: Gilels, Gulda,

Ashkenazy, Arrau, Schnabel, Haskil, Berman, Kempff, Backhaus, Gieseking, Rosen, Frank, Perahia, Bishop, Brendel, Gould, Davidovich, Solomon, Rubinstein. Performances with missing repeats and/or poor sound quality tended to occur towards the end.[8]

C. Analysis

The raw data constituted a 9 (subjects) by 20 (performances) by 20 (adjective pairs) matrix of numerical ratings. Since there was considerable variability between judges' ratings, each adjective scale was first examined as to whether there was any consistency among subjects at all. The criterion was that Kendall's coefficient of concordance be significant ($p < .05$). Three adjective scales (serious/playful, effortful/facile, realistic/idealistic) did not meet this criterion and were eliminated. Since two performances, by the computer and by Gould, often elicited the most extreme ratings, the criterion was applied once more with these two performances omitted. Five additional scales (expressive/inexpressive, cold/warm, spontaneous/deliberate, coherent/incoherent, rigid/flexible) were eliminated at this stage. This left twelve scales for further analysis. The highest coefficient of concordance for all 20 performances (0.59, $p < .0001$) was shown, not surprisingly, by fast/slow, followed by three scales (smooth/rough, excessive/restrained, Beethovenian/UnBeethovenian) with coefficients of 0.33-0.34 ($p < .0001$); the remainder had low but still significant Kendall coefficients. It was gratifying to find that the important Beethovenian/Un-Beethovenian scale was used with some consistency by the subjects.

The data were subsequently averaged across the nine subjects' judgments, which resulted in a 20 by 12 data matrix. The 12 × 12 intercorrelation matrix was subjected to a principal components analysis with Varimax rotation, to reduce the 12 rating scales to a smaller number of evaluative dimensions.

D. Results

1. Factor structure of rating scales

The analysis yielded four factors with eigenvalues larger than 1; together, they accounted for 88 percent of the variance in the data. After Varimax rotation, that variance was divided fairly equally between the four factors. The factor loadings of the twelve rating scales (rearranged) are shown in Table 7; loadings smaller than 0.25 have been suppressed for the sake of clarity. The polarities of Factors 3 and 4 have been reversed for easier labeling.

Table 7. *Sorted rotated factor loadings. Positive loadings represent the second adjective in a pair. Loadings smaller than +/-0.25 have been omitted.*

	Factor			
	1	2	3	4
Soft/hard	0.935			
Relaxed/tense	0.900			
Smooth/rough	0.701	0.463	-0.356	-0.348
Excessive/restrained	-0.349	-0.830		
Consistent/variable		0.814		0.410
Usual/unusual		0.730	-0.372	-0.415
Sloppy/precise	-0.479	-0.701	0.279	
Superficial/deep	-0.268		0.900	
Strong/weak	-0.273		-0.887	
Beet/Un-Beethovenian	0.257		-0.771	-0.444
Fast/slow				-0.925
Pessimistic/optimistic				0.842

The factors can be interpreted without much difficulty. Factor 1 has its highest loadings on "hard," "tense," and "rough," as opposed to soft, relaxed, and smooth. It will be called *Force*. Factor 2 loads highly on "excessive," "variable," "unusual," and "sloppy," as opposed to restrained, consistent, usual, and precise. It will be dubbed *Individuality*. Factor 3 is characterized by the attributes "deep," "strong," and "Beethovenian," as opposed to superficial, weak, and un-Beethovenian. Interestingly, this factor reveals depth and strength (but not restraint) as the primary correlates of the subjects' idea of "Beethovenian." It will be called *Depth*. Finally, Factor 4 loads highly on "fast" and "optimistic," as opposed to slow and pessimistic. Clearly, it is a tempo factor, and it is interesting that the listeners associated optimism so strongly with a fast tempo. It will be called *Speed*. Note that only Depth has a simple relationship to positive/negative evaluation or preference; Force, Individuality, and Speed most likely have a curvilinear relationship, with neither extreme being desirable.

2. Factor scores of performances

Let us examine now how the individual performances ranked on these evaluative dimensions, and what characteristics of the performances or of the performers might be responsible for these rankings. The factor scores are shown in Table 8. On the Force factor, the highest scores were exhibited by Gould, Berman, and Brendel; the lowest scores, by Rubinstein, Arrau, Gilels, and Kempff. It is conceivable that

this factor was influenced by the relative loudness and sonic quality (e.g., "harshness") of the recordings, which were not controlled in any way; however, the author, having taken notes about these aspects of the recordings, sees no obvious relationship. More likely, some acoustic correlate of the pianists' "touch" was involved, such as their degree of legato playing or amplitude dynamics, which were not measured in the present study. Interestingly, a loose relationship with the artists' age is suggested, older artists tending to have negative scores (i.e., less Force). Also, the two women (Davidovich, Haskil) have moderately negative Force scores. To the extent that extremes are to be avoided along the Force dimension, Perahia and Schnabel obtained the most desirable scores on this factor (close to zero).

Table 8. *Factor scores of the 20 performances.*

	Factor			
	1	2	3	4
Arrau	-1.385	0.280	-0.657	-2.199
Ashkenazy	-0.283	-0.481	-0.873	0.563
Backhaus	0.511	2.014	-1.024	1.341
Berman	1.719	0.251	0.680	-0.325
Bishop	0.662	-0.725	0.507	1.254
Brendel	1.172	-1.050	0.479	-0.075
Davidovich	-0.559	-0.253	0.657	0.194
Frank	-0.658	0.002	1.332	0.576
Gieseking	-0.103	-0.657	-2.290	1.157
Gilels	-1.318	-0.204	0.305	-1.378
Gould	2.262	1.472	-0.313	-1.727
Gulda	-0.413	-0.654	0.291	1.261
Haskil	-0.418	-0.123	0.506	-0.162
Kempff	-1.082	0.147	-0.783	-0.168
Perahia	0.033	-0.300	1.800	0.245
Rosen	0.348	0.220	0.460	-0.728
Rubinstein	-1.503	1.449	0.517	0.131
Schnabel	0.060	1.677	-0.285	0.943
Solomon	0.523	-1.426	0.479	-0.115
Computer	0.433	-1.639	-1.785	-0.789

The highest scores by far on the Individuality factor were shown by Backhaus, Schnabel, Gould, and Rubinstein; the lowest scores, by the computer performance, Solomon, and Brendel, followed by Bishop, Gieseking, and Gulda. Several of these performances with negative scores were relatively deadpan (computer, Solomon, Gieseking); the others were probably without highly distinctive properties. The four most unconventional performances, on the other hand, were indeed so: Backhaus introduced striking tempo changes, Schnabel used quirky articulation, Gould was unusually slow and plodding, and Rubinstein used exaggerated expression. Either extreme was avoided most effectively by Frank, Haskil, and Kempff, who scored in the middle range. Given the many different ways in which a performance can be unconventional, it seems unlikely that a single physical correlate of this dimension could be found.

The Depth factor is of special interest here because it represents the listeners' concept of "Beethovenian." The highest score was obtained by Perahia, followed by Frank; these two scores were far ahead of the rest. The lowest scores were shown by Gieseking and the computer performance. Since the latter exhibited Clynes' Beethoven pulse most clearly, it is apparent that Depth scores do not have a positive relationship with the presence of such a pulse.

Finally, the Speed factor clearly contrasted fast performances (Backhaus, Gulda, Bishop) with slow ones (Arrau, Gould, Gilels). Here, indeed, there was a straightforward physical correlate: The correlation of the factor scores with the computed metronome speeds (Table 2, column B) was 0.78 ($p < .001$), just slightly below the correlation between the average ratings on the fast/slow scale itself and the metronome speeds (0.83).

In an attempt to identify possible correlates of the four evaluative dimensions in the timing patterns of the performances, correlations were computed between the factor scores just discussed (Table 8) and the factor loadings of the 20 performances in the earlier analysis of the timing data (Table 4). It should be kept in mind that the timing data derived from the Minuet only, while the evaluations were based not only on the complete performances, but also on many other aspects besides timing. Nevertheless, there were several significant correlations. Individuality correlated negatively (-0.61, $p < .01$) with Timing Factor 3, which represented the V-shaped pulse; it will be recalled that the computer performance and several other deadpan performances ranked lowest on Individuality, which is quite reasonable. Depth correlated positively (0.62, $p < .01$) with Timing Factor 1, which represented mainly the marking of major phrase boundaries, but not (-0.14) with Timing Factor 3 (the Beethoven pulse). Finally, Speed correlated positively (0.63, $p < .01$) with Timing Factor 2, which represented expressive deviations in Bars 1 and 6 as well as a faster tempo for the second half of the Minuet. It also correlated negatively (-0.45, $p < .05$) with

Timing Factor 1, suggesting that the slower performances tended to emphasize phrase boundaries more, whereas the faster performances tended to focus more on certain other expressive deviations. Force did not correlate significantly with any of the three timing factors.

3. Performance styles

The data were analyzed in yet another way, by transposing the 12 by 20 matrix of judgments and conducting another principal components analysis. In this case, the correlations were computed across the 12 rating scales and represented the similarities between the rating "profiles" of the 20 performances. (Because there were only 12 scales, the 20×20 correlation matrix was singular and of rank 11.) Five factors (with eigenvalues greater than 1) accounted for 91 percent of the variance. After Varimax rotation, the first factor accounted for 38 percent of the variance, with the remainder divided about equally among the other four factors. The factor loadings of the 20 performances are shown in Table 9.

One way of interpreting this structure is that, by means of 12 rating scales (or four evaluative dimensions), the judges were able to distinguish five "performance styles." The first factor seemed to reflect a general performance standard; 10 pianists (Frank, Davidovich, Perahia, Gulda, Bishop, Haskil, Solomon, Brendel, Ashkenazy, Rubinstein) had high positive loadings, and only one (Gould) had a negative loading. The other four factors seemed to reflect more individual interpretive styles: Factor 2 was represented primarily by Schnabel and Backhaus; Factor 3 by Berman, Rosen, Gould, and Brendel; Factor 4 by Arrau, Gilels, and Kempff; and Factor 5 by Gieseking, the computer, and Ashkenazy. (Note that only two pianists, Brendel and Ashkenazy, have relatively high loadings on more than one factor.) Comparison with Table 8 reveals that Factors 2 and 5 reflect high and low Individuality respectively; Factors 3 and 4 reflect high and low Force, respectively; and Factor 1 represents the "middle of the road" performances. Note, however, that all factors were strictly orthogonal. This might be interpreted as indicating that the adjective pairs that defined Individuality and Force did not constitute true polar opposites but different dimensions. This is quite plausible in the case of Individuality, which really represents deviations from the norm in different directions and by different means.

Table 9. *Sorted rotated factor loadings of the 20 performances on five "performance style" factors that emerged from the principal components analysis of their rating "profiles." Loadings smaller than +/-0.25 have been omitted.*

	Factor				
	1	2	3	4	5
Frank	0.944				
Davidovich	0.941				
Perahia	0.932				
Gulda	0.892				0.307
Bishop	0.858				0.279
Haskil	0.815	-0.456			
Solomon	0.789	-0.437			0.296
Brendel	0.642	-0.290	0.623		
Rubinstein	0.598	0.333	-0.283	0.473	-0.358
Schnabel		0.945			
Backhaus		0.880		-0.305	
Berman			0.888		
Rosen	0.397		0.677	0.291	
Gould	0.627	0.323	0.660		
Arrau				0.918	
Gilels	0.439			0.517	
Kempff	0.398		-0.474	0.616	0.330
Gieseking			-0.308		0.905
Computer		-0.438		0.301	0.773
Ashkenazy	0.610				0.619

V. GENERAL DISCUSSION

The present study addressed three broad issues: (1) the presence or absence of a "Beethoven pulse" in human performances; (2) general characteristics of expressive timing patterns in different expert performances of the same piece; and (3) listeners' evaluation of performances. These topics will be discussed in turn.

A. The search for the Beethoven pulse

One motivation of the present study was to search for a pulse-like timing pattern in human performances of a Beethoven piece, similar to that discovered subjectively by Clynes (1983) and implemented in his computer performance of the same piece. Before summarizing the outcome of that search, some limitations and strengths of the study should be pointed out.

One severe limitation is obviously that only a single composition was examined. It could be that the composition chosen is not typical of Beethoven or that its Minuet-like character dominated specifically Beethovenian characteristics. In the author's opinion, however, the piece is quite characteristic and not very Minuet-like, and its choice was not inappropriate. A more serious problem of restricting the investigation to a single composer is that any timing patterns found, even if they resemble the Beethoven pulse, may not be specific to Beethoven but may be general features of music performance.

A second limitation is that only timing patterns were examined. It could be that accent (amplitude) patterns interact with timing variations, so that considering the timing pattern alone might present a distorted picture. Nevertheless, an examination of the timing component by itself is a defensible methodological strategy, for which there are numerous precedents in the literature. Of course, no conclusions are warranted with regard to the amplitude component of Clynes' Beethoven pulse, which may or may not have been present in the performances examined here.

Third, the present study focused primarily on the "higher-level" pulse, at the level of quarter-notes, because of the paucity of sixteenth-notes in the piece. Nevertheless, some information relevant to the "basic," lower-level pulse was also obtained.

Balanced against these limitations are a number of strengths of the present research. The study employed a large sample of performances and obtained a large number of measurements from each. The precision of measurement was more than sufficient for its purpose. The artists represent many of the finest interpreters of Beethoven's music in this century. The data were subjected to rigorous statistical analysis. Thus, within the limitations stated above, the search for a Beethoven pulse was fairly exhaustive. Finally, results from an earlier perceptual study (Repp, 1989) indicated that Clynes' Beethoven pulse improved the computer performance of the piece chosen; this provided a valid basis for expecting similar patterns in human performances.

Was a Beethoven pulse found? Certainly, none of the 19 human performances showed a timing pattern closely resembling that of the computer performance. More specifically, none of the human performances showed any relatively constant, pulse-like timing pattern; rather, the timing pattern varied from bar to bar according to musical demands. Thus, in general, human performances not only did not show the specific Beethoven pulse "discovered" by Clynes; they did not show *any* constant pulse.

Consistent with this observation is the finding that the computer performance, even though it had been rated favorably by listeners in comparison to a deadpan performance (Repp, 1989), did not fare well in comparison with human performances. It received the lowest rating on the "Beethovenian" scale, even lower than Gould's leaden performance; in terms of the evaluation factors, it scored lowest on Individuality and second-lowest on Depth. Clearly, a pulse-like timing pattern is not very favorably received by listeners when the alternatives are musically varied timing patterns.

Despite a number of manual adjustments by Clynes, the computer performance did not exhibit the full richness of a human performance; it was still an artificial performance generated for the purpose of testing the effectiveness of a single, isolated microstructure component. Since there are so many other sources of expressive variation that perturb the surface timing pattern of human performances, the composer's pulse may be "underlying," hidden, or intermittent. The way in which an underlying pulse might combine with other sources of timing variation, or how a listener might perceive a constant pulse through a pattern of surface variability, are issues that Clynes has not discussed explicitly. The present data are not incompatible with the presence of an underlying V-shaped pulse pattern in some of the human

performances of the Minuet. The origin and interpretation of that pattern remain uncertain, however. For example, it may just as well derive from structural musical factors than from an underlying pulse: The harmonic and melodic structure of the Minuet tended to weight the first and third quarter-notes in a bar more heavily than the second. Moreover, there may well be a general tendency of performers to shorten the second of three quarter-notes, which is not specific to a particular piece or composer. Indeed, Gabrielsson et al. (1983) found lengthening of either the first or the third beat, but never of the second beat, in musicians' performances of folktunes in 3/4 time.

Even if there was some underlying pulse, it apparently contributed little to the impression of a performance on listeners. That impression, to the extent that it derived from the timing pattern, was governed primarily by pianists' temporal marking of major and minor phrase boundaries, and of one prominent melodic excursion (bar 6).

Similar conclusions apply to the basic, lower-level pulse, even though there were less data available. The relative timing of sixteenth- and eighth-notes was clearly dependent on the musical context, and no pulse-like constancy was evident. In the Trio and the Coda, the observed timing patterns of sixteenth-notes strongly contradicted the ratios specified by Clynes.

The fact that the composer's pulse, as conceived by Clynes, is inherently insensitive to the musical structure of a specific piece seems to preordain a minor role for it, if any, in music performance and evaluation. Since expressive variations based on structural characteristics of the music are large and almost ubiquitous, there is little room for an autonomous pulse to come to the fore. Wolff (1979, p. 15), transmitting Artur Schnabel's views, has made this pertinent comment: "The term 'recreation' has often given rise to the misunderstanding that the interpreter can attempt a revival of the personality of the composer at the moment of creation. The futility of such attitudes is generally acknowledged. We know that all interpretive re-creation depends on the awareness of the structure and objective character of a composition." The presence of a composer's personal pulse in expert performances remains to be demonstrated.

B. Expressive timing in performance

Three general observations can be made from the present data. First, far from being idiosyncratic, the timing patterns of individual artists' performances largely adhere to a common standard. This was demonstrated by the fact that, in the analysis of the Minuet timing patterns, the first two principal components accounted for 71% of the variance. One of these components represents primarily the lengthening at phrase boundaries (a well-known phenomenon; see, e.g., Clarke, 1988; Shaffer & Todd, 1987; Todd, 1985), whereas the other reflects several other types of expressive variation, such as lengthening of salient melodic inflections and tempo changes within and between sections. That the timing variation appears to be governed by two independent dimensions is a finding worth following up in future research. Individual variations consist primarily in the extent to which the structural markers captured by these two dimensions are applied. There seems to be relatively little room for truly idiosyncratic variation in relative timing, at least in the present Minuet, but overall tempo, accent patterns, and articulation offer many additional degrees of freedom to the individual performer.

Second, it is evident that the timing pattern is reproduced with a high degree of precision across repetitions of the same music. This was already noted by Seashore (1938), as well as by others in more recent research (e.g., Gabrielsson, 1987; Palmer, 1989a; Shaffer & Todd, 1987). Although some compositions may call for subtle variations between repeats, in the present Minuet there was no evidence of any systematic timing changes, except at the beginnings and ends of sections, where there was either an actual change in the music or the distinction between continuity and finality had to be conveyed.

Third, the timing patterns at all levels are dependent on the musical structure. Thus, not only the timing of quarter-notes but also that of eighth- and sixteenth-notes varied substantially with their musical function. For example, sixteenth-notes following dotted eighth-notes were generally prolonged in the Minuet, where they were part of an upbeat, but generally shortened in the Trio, where they fell on the downbeat. The same rhythmic pattern can be performed with very different temporal modulations, according to musical requirements (cf. Gabrielsson et al., 1983).[9]

C. Performance evaluation

The systematic description and evaluation of different performances of the same music has

been studied relatively little by psychologists, compared to the considerable literature on listeners' reactions to different compositions. Yet, music critics, jurors at competitions, and discriminating music lovers engage continuously in such judgments which, despite a considerable amount of subjectivity, are by no means totally idiosyncratic. The present results, though they are very preliminary, do suggest that there is some consistency among judges; moreover, part of the variability may well result from lack of skill in using rating scales rather than from genuine judgmental differences.

The Individuality factor obtained in the analysis of the rating scales, as well as the "middle-of-the-road" general factor obtained in the analysis of the performances-as-judged, suggest that musically experienced listeners refer to similar internal performance standards. This common standard is most likely one that includes the basic expressive variations required by the musical structure, without which a performance would be perceived as atypical, impoverished, and unmusical. While there is rarely a single definitive performance of a given piece of music, there can well be a typical or average performance that listeners agree on. It is with respect to their evaluation of deviations from this common standard that listeners show differences of opinion. If the deviations are gross, their reactions may be uniformly negative. (An example is Gould's performance in the present set, though its deviations, apart from its slow tempo, were mostly in aspects other than timing.) However, if the deviations are imaginative, listeners' evaluative judgments may diverge considerably. (Examples in the present set are the performances by Schnabel and Backhaus.) While such performances are stimulating and provide food for discussion, those that come close to the listener's internal standard simply "sound right," and the listener resonates to them. (Perahia and Frank perhaps came closest to that ideal.)

While the Individuality dimension obtained in the present analysis implies reference to a conventional standard and thus may be peculiar to the evaluation of cultural artifacts, the other three dimensions—Force, Depth, and Speed—are clearly related to the three traditional dimensions of the semantic differential—Potency, Evaluation, and Activity—which have also been obtained in studies employing varied musical materials (see, e.g., de la Motte-Haber, 1985). This result is very encouraging, as it suggests that effective rating scales for the formal comparison and evaluation of musical performances could be developed.

REFERENCES

Becking, G. (1928). *Der Musikalische Rhythmus als Erkenntnisquelle.* Augsburg: Filser.

Bengtsson, I., & Gabrielsson, A. (1980). Methods for analyzing performance of musical rhythm. *Scandinavian Journal of Psychology, 21,* 257-268.

Bruhn, H. (1985). Traditionelle Methoden der Musikbeschreibung. In H. Bruhn, R. Oerter, & H. Roesing (Eds.), *Musikpsychologie. Ein Handbuch in Schlüsselbegriffen* (pp. 494-501). Munich: Urban & Schwarzenberg.

Clarke, E. F. (1982). Timing in the performance of Erik Satie's 'Vexations.' *Acta Psychologica, 50,* 1-19.

Clarke, E. F. (1988). Generative principles in music performance. In J. A. Sloboda (Ed.), *Generative processes in music* (pp. 1-26). Oxford: Clarendon Press.

Clynes, M. (1969). Toward a theory of man: Precision of essentic form in living communication. In K. Leibovic & J. Eccles (Eds.), *Information processing in the nervous system* (pp. 177-206). New York: Springer-Verlag.

Clynes, M. (1983). Expressive microstructure in music, linked to living qualities. In J. Sundberg (Ed.), *Studies of music performance* (pp. 76-181). Stockholm: Royal Academy of Music.

Clynes, M. (1986). Generative principles of musical thought: Integration of microstructure with structure. *Communication and Cognition AI, 3,* 185-223.

Clynes, M. (1987a). What can a musician learn about music performance from newly discovered microstructure principles (PM and PAS)? In A. Gabrielsson (Ed.), *Action and perception in rhythm and music* (pp. 201-234). Stockholm: Royal Academy of Music.

Clynes, M. (1987b). *Computerized system for imparting an expressive microstructure to succession of notes in a musical score.* (United States Patent No. 4,704,682).

de la Motte-Haber, H. (1985). *Handbuch der Musikpsychologie.* Laaber, FRG: Laaber-Verlag.

Gabrielsson, A. (1974). Performance of rhythm patterns. *Scandinavian Journal of Psychology, 15,* 63-72.

Gabrielsson, A. (1987). Once again: The theme from Mozart's Piano Sonata in A Major (K.331). In A. Gabrielsson (Ed.), *Action and perception in rhythm and music* (pp. 81-104). Stockholm: Royal Academy of Music.

Gabrielsson, A., Bengtsson, I., & Gabrielsson, B. (1983). Performance of musical rhythm in 3/4 and 6/8 meter. *Scandinavian Journal of Psychology, 24,* 193-213.

Hartmann, A. (1932). Untersuchungen über metrisches Verhalten in musikalischen Interpretationsvarianten. *Archiv für die gesamte Psychologie, 84,* 103-192.

Hevner, K. (1936). Experimental studies of the elements of expression in music. *American Journal of Psychology, 48,* 246-268.

Kaiser, J. (1975). *Beethovens 32 Klaviersonaten und ihre Interpreten.* Frankfurt: S. Fischer.

Lindblom, B. (1978). Final lengthening in speech and music. In E. Garding, G. Bruce, & R. Bannert (Eds.), *Nordic prosody* (pp. 85-100). Sweden: Department of Linguistics, Lund University.

Palmer, C. (1989). Mapping musical thought to musical performance. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 331-346.

Povel, D. J. (1977). Temporal structure of performed music: Some preliminary observations. *Acta Psychologica, 41,* 309-320.

Repp, B. H. (1989). Expressive microstructure in music: A preliminary perceptual assessment of four composers' 'pulses.' *Music Perception, 6,* 243-273.

Repp, B. H. (1990). Further perceptual evaluations of 'composers' pulses' in computer performances of classical piano music. *Music Perception, 8,* 1-33.

Seashore, C. E. (1938). *Psychology of music*. New York: McGraw-Hill.

Senju, M., & Ohgushi, K. (1987). How are the player's ideas conveyed to the audience? *Music Perception, 4*, 311-324.

Shaffer, L. H. (1981). Performances of Chopin, Bach, and Bartok: Studies in motor programming. *Cognitive Psychology, 13*, 326-376.

Shaffer, L. H. (1984). Timing in solo and duet piano performances. *Quarterly Journal of Experimental Psychology, 36A*, 577-595.

Shaffer, L. H., Clarke, E. F., & Todd, N. P. (1985). Metre and rhythm in piano playing. *Cognition, 20*, 61-77.

Shaffer, L. H., & Todd, N. P. (1987). The interpretive component in musical performance. In A. Gabrielsson (Ed.), *Action and perception in rhythm and music* (pp. 139-152). Stockholm: Royal Academy of Music.

Sloboda, J. A. (1983). The communication of musical metre in piano performance. *Quarterly Journal of Experimental Psychology, 35A*, 377-396.

Sloboda, J. A. (1985). Expressive skill in two pianists: Metrical communication in real and simulated performances. *Canadian Journal of Psychology, 39*, 273-293.

Sundberg, J. (1988). Computer synthesis of music performance. In J. A. Sloboda (Ed.), *Generative processes in music* (pp. 52-69). Oxford: Clarendon Press.

Thompson, W. F. (1989). Composer-specific aspects of musical performance: An evaluation of Clynes's theory of pulse for performances of Mozart and Beethoven. *Music Perception, 7*, 15-42.

Todd, N. (1985). A model of expressive timing in tonal music. *Music Perception, 3*, 33-58.

Wolff, K. (1979). *Schnabel's interpretation of piano music*. London: Faber Music.

FOOTNOTES

Journal of the Acoustical Society of America, 88, 622-641 (1990).

[1]A few words about the artists may be in order for readers not acquainted with them. The sample includes two women (Davidovich, Haskil) and 17 men. Nine pianists (Backhaus, Gieseking, Gilels, Gould, Haskil, Kempff, Rubinstein, Schnabel, Solomon) are dead; the others are still active on the world's concert stages at the time of writing. Schnabel is universally considered an authoritative interpreter of Beethoven's sonatas; Backhaus, Brendel, and Kempff are generally considered Beethoven specialists as well. Other pianists who have played much Beethoven and have recorded the complete sonatas include Arrau, Frank, and Gulda. Since almost every major pianist plays Beethoven sonatas frequently, all of the remaining artists certainly have much experience as Beethoven interpreters, though some are known better from other repertoire. Thus, Gould is known primarily for his Bach recordings and often eccentric performances; Haskil is considered a Mozart specialist; Berman is associated with Romantic virtuoso pieces; Gieseking and Rubinstein, though they frequently played Beethoven, are generally not considered ideal interpreters of his music (the former being too fleet, the latter too effusive). The different ages of the artists at the presumed time of recording may also be noted; they range from rather young (Bishop, Gulda, Perahia) to rather old (Arrau, Backhaus, Kempff, Rubinstein). Finally, the artists may also be divided into three major national groups: German-Austrian (Backhaus, Brendel, Gieseking, Gulda, Haskil, Kempff, Schnabel, as well as the Chilean-born but German-educated Arrau), Russian (Ashkenazy, Berman, Davidovich, Gilels), and American-Canadian (Bishop, Frank [of German origin], Gould, Perahia, Rosen); this leaves only the British pianist Solomon, and the very cosmopolitan Polish-born Rubinstein. It will be of interest to see whether any of these characteristics are related to expressive timing patterns in music performance.

[2]To guard against variations in speed caused by the multiple transfers of the recordings using a variety of playback equipment, the sound wave corresponding to the initial note of each performance (a B-flat) was subjected to spectral analysis. The FFT spectrum was calculated over a 102.4 ms Hamming window placed roughly over the center of the waveform, and the frequency of the lowest harmonic was determined with a resolution of 4 Hz. Each of the 20 recordings, including the computer performance from Clynes' laboratory, yielded a fundamental frequency of 244 Hz plus/minus 4 Hz (3 plus, 4 minus), so that the recording/playback speeds may be considered comparable. The average frequency of the first note, however, was higher than expected. B-flat, being one semitone above the A one octave below the standard A of 440 Hz, should be 5.9 percent higher than 220 Hz, or at about 233 Hz. Thus it seems that all recordings were played somewhat fast, almost one semitone too high. However, since this difference could not be traced directly to any piece of equipment, no correction was made in the measurements.

[3]It is interesting to compare these metronome values to various opinions about what the tempo *should* be. Urtext editions of Beethoven's piano sonatas (Breitkopf & Hartel, Kalmus, Peters) do not have metronome markings. A perusal of a large number of other editions, however, revealed several in which the editors had inserted their own suggested metronome speeds. For the Minuet, they range from 84 qpm (d'Albert) to 88 qpm (Bülow/Lebert) to 96 qpm (Epstein, Schnabel) to 104-108 qpm (Casella). Schnabel, in addition, gives separate, faster metronome indications for the Trio (108 qpm) and the Coda (100 qpm), whereas Casella indicates that the Trio should be taken slower (96 qpm) than the Minuet. These metronome speeds are quite fast compared with the present performances. Only the slowest marking, by d'Albert, is close to the average speed taken by this group of pianists. A number of pianists, including Schnabel, play the Trio faster than the Minuet, as suggested in the Schnabel edition. However, Schnabel's tempi fall short of his own recommendations; only Backhaus comes close to those.

[4]Actually, the specification was [105, 88, 107], but the deviations were "attenuated" by 50%, according to Clynes' musical judgment in generating the computer performance. Note that the possibility of varying the modulation depth of a pulse defines a pulse family, rather than a single fixed pattern, for a given composer. However, Clynes usually applies attenuation only to the higher-level (slower) pulse.

[5]All 19 performances were included in this analysis. In the four performances that were missing the fourth repeat, the data of the second repeat were duplicated.

[6]The two-beat pulse pattern is obtained from the four-beat values "by adding the duration of tones 1 and 2, and of tones 3 and 4, respectively, to obtain the duration proportions" (Clynes 1983, p. 161). If applied to Clynes' four-beat pulse, this yields [97.5, 103.5] due to the fact that percentage deviations above and below 100 are not perfectly balanced in Clynes' four-beat specification, perhaps due to omission of decimals. Since the percent increase of one eighth-note OOI in a pair must equal the decrease in the other, the two-beat specification was adjusted to [97, 103]. Clynes' (1983, p. 162) specification of the two-beat Beethoven pulse as [97.5, 100.3] appears to be a mistake.

[7]For one subject (the author's mother) the adjectives were translated into their German equivalents. No particular model was followed in their selection. Though similar adjective pairs have been used in the past to evaluate the expression of different musical compositions (e.g., Hevner, 1936; see Bruhn, 1985), at the time the author was not aware of any scales constructed

specifically for the evaluation of different performances of one and the same piece. Meanwhile, an article by Senju and Ohgushi (1987) has described such a scale containing 15 adjective pairs (translated from the Japanese).

[8]Since experienced listeners were expected to judge the performances against a well-established internal standard, the fixed order was not considered a serious problem. However, counterbalancing should be employed in future, more extensive research.

[9]Many more specific observations could be made in the present data, which may be of interest to students of music performance. For those interested in pursuing such details, the author will be happy to provide the raw data or graphs of individual performances.

Appendix

SR #	Report Date	DTIC #	ERIC #
SR-21/22	January-June 1970	AD 719382	ED 044-679
SR-23	July-September 1970	AD 723586	ED 052-654
SR-24	October-December 1970	AD 727616	ED 052-653
SR-25/26	January-June 1971	AD 730013	ED 056-560
SR-27	July-September 1971	AD 749339	ED 071-533
SR-28	October-December 1971	AD 742140	ED 061-837
SR-29/30	January-June 1972	AD 750001	ED 071-484
SR-31/32	July-December 1972	AD 757954	ED 077-285
SR-33	January-March 1973	AD 762373	ED 081-263
SR-34	April-June 1973	AD 766178	ED 081-295
SR-35/36	July-December 1973	AD 774799	ED 094-444
SR-37/38	January-June 1974	AD 783548	ED 094-445
SR-39/40	July-December 1974	AD A007342	ED 102-633
SR-41	January-March 1975	AD A013325	ED 109-722
SR-42/43	April-September 1975	AD A018369	ED 117-770
SR-44	October-December 1975	AD A023059	ED 119-273
SR-45/46	January-June 1976	AD A026196	ED 123-678
SR-47	July-September 1976	AD A031789	ED 128-870
SR-48	October-December 1976	AD A036735	ED 135-028
SR-49	January-March 1977	AD A041460	ED 141-864
SR-50	April-June 1977	AD A044820	ED 144-138
SR-51/52	July-December 1977	AD A049215	ED 147-892
SR-53	January-March 1978	AD A055853	ED 155-760
SR-54	April-June 1978	AD A067070	ED 161-096
SR-55/56	July-December 1978	AD A065575	ED 166-757
SR-57	January-March 1979	AD A083179	ED 170-823
SR-58	April-June 1979	AD A077663	ED 178-967
SR-59/60	July-December 1979	AD A082034	ED 181-525
SR-61	January-March 1980	AD A085320	ED 185-636
SR-62	April-June 1980	AD A095062	ED 196-099
SR-63/64	July-December 1980	AD A095860	ED 197-416
SR-65	January-March 1981	AD A099958	ED 201-022
SR-66	April-June 1981	AD A105090	ED 206-038
SR-67/68	July-December 1981	AD A111385	ED 212-010
SR-69	January-March 1982	AD A120819	ED 214-226
SR-70	April-June 1982	AD A119426	ED 219-834
SR-71/72	July-December 1982	AD A124596	ED 225-212
SR-73	January-March 1983	AD A129713	ED 229-816
SR-74/75	April-September 1983	AD A136416	ED 236-753
SR-76	October-December 1983	AD A140176	ED 241-973
SR-77/78	January-June 1984	AD A145585	ED 247-626
SR-79/80	July-December 1984	AD A151035	ED 252-90
SR-81	January-March 1985	AD A156294	ED 257-159
SR-82/83	April-September 1985	AD A165084	ED 266-508
SR-84	October-December 1985	AD A168819	ED 270-831
SR-85	January-March 1986	AD A173677	ED 274-022
SR-86/87	April-September 1986	AD A176816	ED 278-066
SR-88	October-December 1986	PB 88-244256	ED 282-278

SR-89/90	January-June 1987	PB 88-244314	ED 285-228
SR-91	July-September 1987	AD A192081	**
SR-92	October-December 1987	PB 88-246798	**
SR-93/94	January-June 1988	PB 89-108765	**
SR-95/96	July-December 1988	PB 89-155329/AS	
SR-97/98	January-July 1989	PB 90-121161/AS	ED321317
SR-99/100	July-December 1989	PB 90-226143/AS	ED321318
SR-101/102	January-June 1990	PB 91-138479	
SR-103/104	July-December 1990	PB 91-172924	
SR-105/106	January-June 1991		

AD numbers may be ordered from:

U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Road
Springfield, VA 22151

ED numbers may be ordered from:

ERIC Document Reproduction Service
Computer Microfilm Corporation (CMC)
3900 Wheeler Avenue
Alexandria, VA 22304-5110

In addition, Haskins Laboratories Status Report on Speech Research is abstracted in *Language and Language Behavior Abstracts*, P.O. Box 22206, San Diego, CA 92122

**Accession number not yet assigned

Haskins Laboratories Status Report on Speech Research

SR-105/106
JANUARY-JUNE 1991

Contents

- Phonology and Beginning Reading Revisited
  Isabelle Y. Liberman 1
- The Role of Working Memory in Reading Disability
  Susan Brady 9
- Working Memory and Comprehension of Spoken Sentences: Investigations of Children with Reading Disorder
  Stephen Crain, Donald Shankweiler, Paul Macaruso, and Eva Bar-Shalom 23
- Explaining Failures in Spoken Language Comprehension by Children with Reading Disability
  Stephen Crain and Donald Shankweiler 43
- How Early Phonological Development Might Set the Stage for Phoneme Awareness
  Anne E. Fowler 53
- Modularity and the Effects of Experience
  Alvin M. Liberman and Ignatius G. Mattingly 65
- Modularity and Dissociations in Memory Systems
  Robert G. Crowder 69
- Representation and Reality: Physical Systems and Phonological Structure
  Catherine P. Browman and Louis Goldstein 83
- Young Infants' Perception of Liquid Coarticulatory Influences on Following Stop Consonants
  Carol A. Fowler, Catherine T. Best, and Gerald W. McRoberts 93
- Extracting Dynamic Parameters from Speech Movement Data
  Caroline L. Smith, Catherine P. Browman, Richard S. McGowan, and Bruce Kay 107
- Phonological Underspecification and Speech Motor Organization
  Suzanne E. Boyce, Rena A. Krakow, and Fredericka Bell-Berti 141
- Task Coordination in Human Prehension
  Patrick Haggard 153
- Masking and Stimulus Intensity Effects on Duplex Perception: A Confirmation of the Dissociation Between Speech and Nonspeech Modes
  Shlomo Bentin and Virginia Mann 173
- The Influence of Spectral Prominence on Perceived Vowel Quality
  Patrice Speeter Beddor and Sarah Hawkins 187
- On the Perception of Speech from Time-varying Acoustic Information: Contributions of Amplitude Variation
  Robert E. Remez and Philip E. Rubin 215
- Subject Definition and Selection Criteria for Stuttering Research in Adult Subjects
  Peter J. Alfonso 231
- Vocal Fundamental Frequency Variability in Young Children: A Comment on *Developmental Trends in Vocal Fundamental Frequency of Young Children* by M. Robb and J. Saxman
  Margaret Lahey, Judy Flax, Katherine Harris, and Arthur Boothroyd 243
- Patterns of Expressive Timing in Performances of a Beethoven Minuet by Nineteen Famous Pianists
  Bruno H. Repp 247

Appendix 273

BEST COPY AVAILABLE